Exact Community Recovery over Signed Graphs

Xiaolu Wang[*] **Peng Wang**[†] **Anthony Man-Cho So**[*]

[*]Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong
[†]Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor

Abstract

Signed graphs encode similarity and dissimilarity relationships among different entities with positive and negative edges. In this paper, we study the problem of community recovery over signed graphs generated by the signed stochastic block model (SSBM) with two equal-sized communities. Our approach is based on the maximum likelihood estimation (MLE) of the SSBM. Unlike many existing approaches, our formulation reveals that the positive and negative edges of a signed graph should be treated unequally. We then propose a simple two-stage iterative algorithm for solving the regularized MLE. It is shown that in the logarithmic degree regime, the proposed algorithm can exactly recover the underlying communities in nearly-linear time at the information-theoretic limit. Numerical results on both synthetic and real data are reported to validate and complement our theoretical developments and demonstrate the efficacy of the proposed method.

1 INTRODUCTION

Signed graph is a type of undirected graph whose edges can take positive or negative values. Unlike unsigned graphs whose edge labels can only be positive (i.e., $+1$), signed graphs employ negative edges (i.e., -1) to characterize the dissimilarity or repulsiveness relationships among different entities. Signed graphs are used to model a wide range of real-world objects, such as friendly and antagonistic relations in social psychology

(Harary, 1953; Cartwright and Harary, 1956; Davis, 1967), common and opposite viewpoints in online social networks (Yardi and Boyd, 2010), trust or distrust between users in bitcoin trading platforms (Dittrich and Matz, 2020), gene regulatory networks of bacteria (Fujita et al., 2012), compensatory interaction networks formed during viral evolution (Quadeer et al., 2018), etc. Recent years have witnessed increasing interest in the analysis of signed graphs in a variety of machine learning tasks, such as graph clustering (Kunegis et al., 2010; Chiang et al., 2012; Mercado et al., 2016; Cucuringu et al., 2019; Mercado et al., 2019), node ranking (Shahriari and Jalili, 2014), link prediction (Chiang et al., 2011, 2014; Ye et al., 2013; Le Falher et al., 2017), node classification (Bosch et al., 2018), graph embedding (Yuan et al., 2017; Wang et al., 2017; Derr et al., 2018), etc.

The study of signed graphs stems from the seminal work by Harary (1953) on the social balance theory, where positive edges encode friend relationships and negative edges encode enemy relationships. In this theory, a signed graph is said to be balanced if the network it represents follows the patterns: "a friend of my friend is my friend", "an enemy of my friend is my enemy", and "an enemy of my enemy is my friend". Such graphs are inherently endowed with community structures. More specifically, it is shown in Harary (1953) that a signed graph is balanced if and only if (i) all of its edges are positive, or (ii) its nodes can be partitioned into 2 disjoint communities such that the edges within communities are positive and the edges between communities are negative.

In this paper, we focus on the problem of identifying the hidden communities over signed graphs that are randomly generated by the signed stochastic block model (SSBM). Specifically, given n nodes that are partitioned into two equal-sized communities, a balanced signed graph is formed by randomly connecting the nodes within the same community with positive edges and the nodes across different communities

with negative edges. To make the SSBM more realistic, we allow the presence of noise that yields unbalanced signed graphs, i.e., there can be negative edges within communities and positive edges across communities. In the logarithmic sparsity regime of the SSBM, the information-theoretic limit, which is the necessary and sufficient condition for exact community recovery, is established in the literature (see, e.g., Jog and Loh (2015); Yun and Proutiere (2016)). In this work, we develop a simple and scalable algorithm for exact community recovery in the SSBM at the information-theoretic limit.

1.1 Related Works

Community detection, which is also known as graph clustering, is one of the most important tasks in network analysis. Motivated by the social balance theory, the rough goal of community detection over signed graphs is to find a partition of its nodes so that there are as many as possible positive edges within communities and negative edges across communities.

There are a collection of works on *correlation clustering* for clustering complete signed graphs. The problem is NP-hard while suboptimal partition schemes can be obtained in polynomial time through approximation algorithms. For example, the formulation in Bansal et al. (2004); Giotis and Guruswami (2006) equally weighs the number of positive (resp. negative) edges within clusters and the number of negative (resp. positive) edges across clusters. Later, Coleman et al. (2008) considered correlation clustering with 2 communities based on a local search algorithm. More recently, Puleo and Milenkovic (2015) extended the correlation clustering model to allow for bounds on the sizes of the communities.

Spectral methods are among the most popular clustering methods, which proceeds by computing the eigenvectors of a matrix associated with the graph, followed by k-means clustering. For example, Kunegis et al. (2010) studied the 2-way spectral clustering over signed graphs based on the signed Laplacian matrix. Later, Chiang et al. (2012) extended the signed graph clustering problem to k communities by minimizing the balanced normalized graph cut, which yields an eigen-problem as well. Recently, Bonchi et al. (2019) considered the problem of detecting 2 small conflicting subsets of nodes in a signed graph, which is formulated as the so-called "discrete eigenvector" problem. Chu et al. (2016); Tzeng et al. (2020) generalized the problem to $k \geq 2$ subsets and approximately solved the resulting formulation using a spectral method.

Stochastic block model (SBM) is a generative model for random graphs that admit community structures.

It provides a useful benchmark for validating and comparing different algorithms for community detection (Abbe, 2018). Over the past years, great progress has been made on establishing the fundamental limits for detecting communities in various SBMs (Mossel et al., 2014; Abbe et al., 2015; Abbe and Sandon, 2015; Yun and Proutiere, 2016) and on developing computationally tractable algorithms that can recover the hidden communities at the information-theoretic limits (Hajek et al., 2016; Gao et al., 2017; Amini and Levina, 2018; Wang et al., 2020, 2021a,b). In the context of signed graph clustering, Cucuringu et al. (2019) proposed a formulation based on a generalized eigenproblem and Mercado et al. (2016, 2019) proposed approaches based on the matrix power means Laplacian. They provided recovery guarantees for signed graphs based on SBMs. However, their approaches cannot achieve exact recovery at the information-theoretic limit of the SSBM, while ours is able to do so. Yun and Proutiere (2016) established the information-theoretic bound for a general type of labeled stochastic block model and proposed a spectral partition algorithm that achieves exact recovery in $\mathcal{O}(n \log^2 n)$ time. Similar to the SSBM, the censored block model (CBM) (Abbe et al., 2014) generates graphs with ternary edge labels $\{*, 0, 1\}$, where $*$ encodes the absence of information. Semidefinite programming is employed to recover the ground-truth communities in the CBM, which is solvable in polynomial time but is not scalable.

1.2 Main Contributions

In this work, we study the modeling and algorithmic issues for community detection, or more specifically, community recovery, in the SSBM. Our main contributions are summarized as follows:

(1) We derive the maximum likelihood estimation (MLE) formulation, which is nonconvex and discrete, for community recovery in the SSBM. This formulation reveals that the positive and negative parts of a signed graph should be treated unequally, which is different from many approaches in the literature (see, e.g., Kunegis et al. (2010); Mercado et al. (2016, 2019)).

(2) We propose a simple two-stage iterative algorithm to solve the regularized version of the MLE. Specifically, the first stage employs power iterations (PIs) to obtain an approximate solution. The second stage employs generalized power iterations (GPIs) to iteratively refine the scaled approximate solution. We prove that the proposed algorithm can exactly recover the underlying communities in $\mathcal{O}(n \log^2 n / \log \log n)$ time at the information-theoretic limit.

(3) Since the connectivity parameters of the SSBM are generally unknown, we provide a method to estimate the true parameters with non-asymptotic up-

Xiaolu Wang[*], Peng Wang[†], Anthony Man-Cho So[*]

per bounds on the estimation errors. The estimation method is based on counting the number of edges and triangles in the graph and can be easily implemented. (4) We also conduct experiments on both synthetic and real data. The numerical results support our theoretical developments and demonstrate the efficacy of our proposed approach.

1.3 Notation

We use lower boldface letters, e.g., $\boldsymbol{v}$, to denote vectors. Given a vector $\boldsymbol{v} \in \mathbb{R}^n$, we use v_i to denote its i-th elements, $\|\boldsymbol{v}\|$ to denote its ℓ_2-norm, and $\boldsymbol{v}/|\boldsymbol{v}|$ to denote the vector given by

$$\left(\frac{\boldsymbol{v}}{|\boldsymbol{v}|}\right)_i = \begin{cases} 1, & \text{if } v_i \geq 0, \\ -1, & \text{otherwise,} \end{cases} \quad \text{for } i = 1, 2, \dots, n.$$

We use upper boldface letters, e.g., $\boldsymbol{M}$, to denote matrices. Given a matrix $\boldsymbol{M}$, we use M_{ij} to denote its (i,j)-th element and $\|\boldsymbol{M}\|$ to denote its spectral norm. Given an integer $m \geq 1$, we use $[m]$ to denote the set $\{1, \dots, m\}$. We use $\textbf{Bern}(p)$ to denote the Bernoulli random variable with mean $p \in [0, 1]$. We use $\boldsymbol{I}$, $\boldsymbol{E}$, $\boldsymbol{0}$, and $\boldsymbol{1}$ to denote the identity matrix, the all-one matrix, the all-zero vector, and the all-one vector, respectively, and their dimensions will be clear from the context. Finally, we use "w.p." as the abbreviation for "with probability".

2 PRELIMINARIES AND MAIN RESULTS

In this section, we first formally introduce the SSBM and derive its corresponding MLE formulation. Then, we present the proposed method for exact community recovery based on a regularized version of the MLE. Finally, we give a summary of the recovery guarantee of our method.

2.1 Signed Stochastic Block Model

The SSBM considered in this paper is formally defined as follows:

Definition 1 (Signed Stochastic Block Model)**.** *Let $n \geq 2$ be an integer, $\mathcal{V} := [n]$ be the node set that are partitioned into two equal-sized disjoint communities $\mathcal{V}_1$ and $\mathcal{V}_2$, and $\boldsymbol{x}^* \in \{1, -1\}^n$ be the community label vector such that $x_i^* = 1$ if $i \in \mathcal{V}_1$ and $x_i^* = -1$ if $i \in \mathcal{V}_2$. Let $p^+, p^-, q^+, q^- \in (0, 1]$ be the connectivity probabilities. We say that the signed graph $\mathcal{G}$ is generated according to the signed stochastic block model, denoted by $\textsf{SSBM}(n, \boldsymbol{x}^*, p^+, p^-, q^+, q^-)$, if $\mathcal{G}$ has node set $\mathcal{V}$ and the elements $\{A_{ij}\}_{1 \leq i,j \leq n}$ of its adjacency matrix $\boldsymbol{A}$ are generated independently as follows:*

(i) If nodes i and j belong to the same community and $i \neq j$, then

$$A_{ij} = A_{ji} = \begin{cases} 1, & w.p. \quad p^+, \\ -1, & w.p. \quad p^-, \\ 0, & w.p. \quad 1 - p^+ - p^-. \end{cases} \tag{1}$$

(ii) If nodes i and j belong to different communities, then

$$A_{ij} = A_{ji} = \begin{cases} 1, & w.p. \quad q^+, \\ -1, & w.p. \quad q^-, \\ 0, & w.p. \quad 1 - q^+ - q^-. \end{cases} \tag{2}$$

(iii) There are no self-loops in $\mathcal{G}$, i.e, $A_{ii} = 0$ for all $i \in [n]$.

We should mention that there are some other types of signed SBMs in the literature for analyzing community detection methods over signed graphs. In particular, taking $p^+ = q^-$ and $p^- = q^+$, $\textsf{SSBM}(n, \boldsymbol{x}^*, p^+, p^-, q^+, q^-)$ boils down to the signed SBM in Cucuringu et al. (2019) for 2-way clustering. Besides, Mercado et al. (2016, 2019) considered a slightly different signed SBM that allows simultaneous presence of both positive and negative edges between each pair of nodes.

In this work, our goal is to develop a simple and efficient algorithm that can exactly recover the true community label vector $\boldsymbol{x}^*$ based on the adjacency matrix $\boldsymbol{A}$ generated by $\textsf{SSBM}(n, \boldsymbol{x}^*, p^+, p^-, q^+, q^-)$. We are especially interested in the logarithmic degree regime of the SSBM, i.e.,

$$p^+ = \frac{\alpha^+ \log n}{n}, \ p^- = \frac{\alpha^- \log n}{n}, \tag{3}$$

$$q^+ = \beta^+ \frac{\log n}{n}, \ q^- = \beta^- \frac{\log n}{n}, \tag{4}$$

where $\alpha^+, \beta^+, \alpha^-, \beta^- > 0$. The SSBM in this regime exhibits a sharp information-theoretic threshold. Indeed, it is shown in the literature (Jog and Loh, 2015; Yun and Proutiere, 2016) that it is possible to recover $\boldsymbol{x}^*$ in $\textsf{SSBM}\left(n, \boldsymbol{x}^*, \frac{\alpha^+ \log n}{n}, \frac{\alpha^+ \log n}{n}, \frac{\alpha^+ \log n}{n}, \frac{\alpha^+ \log n}{n}\right)$ with high probability if and only if

$$(\sqrt{\alpha^+} - \sqrt{\beta^+})^2 + (\sqrt{\alpha^-} - \sqrt{\beta^-})^2 \geq 2. \tag{5}$$

Given the above preliminaries, we make the following assumption that will be used in the subsequent discussion:

Assumption 1. *The adjacency matrix $\boldsymbol{A}$ of the signed graph $\mathcal{G}$ is generated according to the $\textsf{SSBM}(n, \boldsymbol{x}^*, p^+, p^-, q^+, q^-)$ given by Definition 1. Moreover, the connectivity probabilities p^+, p^-, q^+, q^- are given by (3) and (4) with $\alpha^+ > \beta^+ > 0$ and $\beta^- > \alpha^- > 0$.*

The requirements $\alpha^+ > \beta^+$ and $\beta^- > \alpha^-$ in Assumption 1 are for ease of technical exposition. Indeed, if $\alpha^+ < \beta^+$ or $\beta^- < \alpha^-$, the main theoretical results in this section still remain valid based on similar technical developments.

2.2 Maximum Likelihood Estimation

Before we proceed, we let $\mathcal{G}^+$ (resp. $\mathcal{G}^-$) be the subgraph formed by the positive (resp. negative) edges of the signed graph $\mathcal{G}$, whose adjacency matrix $\boldsymbol{A}^+$ (resp. $\boldsymbol{A}^-$) is given by

$$A_{ij}^+ = \max\{A_{ij}, 0\} \text{ (resp. } A_{ij}^- = \max\{-A_{ij}, 0\}\text{)}, \quad (6)$$

for all $i, j \in [n]$. Obviously, we have $\boldsymbol{A} = \boldsymbol{A}^+ - \boldsymbol{A}^-$. By Definition 1 and (6), A_{ij}^+ and A_{ij}^- are dependent random variables for all $i, j \in [n]$. Moreover, $\{A_{ij}^+\}_{1 \leq i < j \leq n}$ (resp. $\{A_{ij}^-\}_{1 \leq i < j \leq n}$) are i.i.d. $\mathbf{Bern}(p^+)$ (resp. $\mathbf{Bern}(p^-)$) if nodes i and j belong to the same community and $\{A_{ij}^+\}_{1 \leq i < j \leq n}$ (resp. $\{A_{ij}^-\}_{1 \leq i < j \leq n}$) are $\mathbf{Bern}(q^+)$ (resp. $\mathbf{Bern}(q^-)$) otherwise. Then, we are ready to present the MLE formulation for community recovery in $\mathsf{SSBM}(n, \boldsymbol{x}^*, p^+, p^-, q^+, q^-)$.

Proposition 1. *Suppose that Assumption 1 holds. Then, the maximum likelihood estimator of the ground-truth $\boldsymbol{x}^*$ is the solution to the following problem:*

$$\begin{aligned} \max \quad & \boldsymbol{x}^\top (\mu_n \boldsymbol{A}^+ - \nu_n \boldsymbol{A}^-) \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{x} \in \{1, -1\}^n, \ \mathbf{1}^\top \boldsymbol{x} = 0, \end{aligned} \quad (7)$$

where

$$\mu_n := \log\left(\frac{\alpha^+}{\beta^+}\right) + \log\left(\frac{n - (\beta^+ + \beta^-)\log n}{n - (\alpha^+ + \alpha^-)\log n}\right),$$

$$\nu_n := \log\left(\frac{\beta^-}{\alpha^-}\right) + \log\left(\frac{n - (\alpha^+ + \alpha^-)\log n}{n - (\beta^+ + \beta^-)\log n}\right).$$

The proof of Proposition 3 is deferred to Section A in the supplementary material, from which we note that the MLE formulation (7) seeks a partition of the nodes so that the signed graph $\mathcal{G}$ is as balance as possible, i.e., there are as many as possible positive edges within communities and negative edges across communities. The weight parameters μ_n and ν_n, which depend on the connectivity probabilities, indicate that the positive and negative edges actually contribute unequally to the community recovery task except when $\mu_n = \nu_n$. This is different from many signed graph clustering methods that equally treat the positive and negative edges (see, e.g., Giotis and Guruswami (2006); Coleman et al. (2008); Kunegis et al. (2010); Mercado et al. (2016, 2019)).

2.3 Regularized MLE and Parameter Estimation

Motivated by the MLE formulation in Proposition 1 and the community recovery method for unsigned graphs (Wang et al., 2020), our approach to community recovery over signed graphs is based on the regularized version of Problem (7). Upon defining

$$\xi := \lim_{n \to \infty} \frac{\nu_n}{\mu_n} = \frac{\log(\beta^-/\alpha^-)}{\log(\alpha^+/\beta^+)}, \ \widetilde{\boldsymbol{A}} := \boldsymbol{A}^+ - \xi \boldsymbol{A}^-, \quad (8)$$

we consider the following regularized MLE problem by penalizing the linear constraint $\mathbf{1}^\top \boldsymbol{x} = 0$ in the objective function:

$$\max\left\{\boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x} : \ \boldsymbol{x} \in \{1, -1\}^n\right\}, \quad (9)$$

where $\boldsymbol{W} = \widetilde{\boldsymbol{A}} - \rho \boldsymbol{E}$ and $\rho = \mathbf{1}^\top \widetilde{\boldsymbol{A}} \mathbf{1}/n^2$. Since $\boldsymbol{W}$ is not necessarily positive semi-definite, the objective function of Problem (9) is in general nonconvex. We should point out that the matrix $\boldsymbol{W}$ relies on the weight parameter ξ, which is possibly unknown. In the context of the SSBM, we may only observe the graph $\mathcal{G}$, while the connectivity parameters α^+, β^+, α^-, β^- that defines ξ are unknown. To address this issue, we propose a method to estimate ξ based on $\mathcal{G}$ with a non-asymptotic upper bound on the estimation error. We formally state it in the following proposition.

Proposition 2. *Suppose that Assumption 1 holds. Let N^+ (resp. N^-) be the number of edges, T^+ (resp. T^-) be the number of triangles in $\mathcal{G}^+$ (resp. $\mathcal{G}^-$). We let*

$$\hat{\alpha}^+ = \frac{1}{\log n}\left(\frac{2N^+}{n} + \sqrt[3]{6T^+ - \frac{8N^{+3}}{n^3}}\right), \quad (10\text{a})$$

$$\hat{\beta}^+ = \frac{1}{\log n}\left(\frac{2N^+}{n} - \sqrt[3]{6T^+ - \frac{8N^{+3}}{n^3}}\right), \quad (10\text{b})$$

$$\hat{\alpha}^- = \frac{1}{\log n}\left(\frac{2N^-}{n} + \sqrt[3]{6T^- - \frac{8N^{-3}}{n^3}}\right), \quad (10\text{c})$$

$$\hat{\beta}^- = \frac{1}{\log n}\left(\frac{2N^-}{n} - \sqrt[3]{6T^- - \frac{8N^{-3}}{n^3}}\right), \quad (10\text{d})$$

and

$$\hat{\xi} := \frac{\log(\hat{\beta}^-/\hat{\alpha}^-)}{\log(\hat{\alpha}^+/\hat{\beta}^+)}. \quad (11)$$

Then, for sufficiently large n, it holds with probability at least $1 - n^{-\Omega(1)}$ that

$$|\hat{\xi} - \xi| \leq \kappa \left(\log n\right)^{-\frac{1}{12}},$$

where $\kappa > 0$ is a constant.

Xiaolu Wang[*], **Peng Wang**[†], **Anthony Man-Cho So**[*]

Algorithm 1 Algorithm for Solving Problem (9)

1: **Input:** Signed graph $\mathcal{G}$ with adjacency matrices $\boldsymbol{A}^+, \boldsymbol{A}^-$, positive integers T_1, T_2
2: Set $\xi \leftarrow \log\left(\beta^-/\alpha^-\right)/\log\left(\alpha^+/\beta^+\right)$
3: Set $\widetilde{\boldsymbol{A}} \leftarrow \boldsymbol{A}^+ - \xi \boldsymbol{A}^-$ and $\rho \leftarrow \mathbf{1}^\top \widetilde{\boldsymbol{A}} \mathbf{1}/n^2$
4: Set $\boldsymbol{W} \leftarrow \widetilde{\boldsymbol{A}} - \rho \boldsymbol{E}$
5: Generate $\boldsymbol{y}^0$ uniformly distributed over the unit sphere
6: **for** $t = 1, 2, \ldots, T_1$ **do**
7:     Set $\boldsymbol{y}^t \leftarrow \boldsymbol{W}\boldsymbol{y}^{t-1}/\|\boldsymbol{W}\boldsymbol{y}^{t-1}\|$
8: **end for**
9: Set $\boldsymbol{x}^0 \leftarrow \sqrt{n}\boldsymbol{y}^{T_1}$
10: **for** $t = 1, 2, \ldots, T_2$ **do**
11:     Set $\boldsymbol{x}^t \leftarrow \boldsymbol{W}\boldsymbol{x}^{t-1}/|\boldsymbol{W}\boldsymbol{x}^{t-1}|$
12: **end for**

The above parameter estimation method requires knowledge of the numbers of edges and triangles in $\mathcal{G}^+$ and $\mathcal{G}^-$, respectively. Indeed, given the adjacency matrix $\boldsymbol{A}^+$ (resp. $\boldsymbol{A}^-$), we can obtain the number of edges in $\mathcal{G}^+$ (resp. $\mathcal{G}^-$) by the formula

$$N^+ = \sum_{1 \le i < j \le n} A_{ij}^+ \ \left(\text{resp. } N^- = \sum_{1 \le i < j \le n} A_{ij}^-\right)$$

and the number of triangles in $\mathcal{G}^+$ (resp. $\mathcal{G}^-$) by the formula

$$T^+ = \frac{1}{6}\operatorname{tr}\left((\boldsymbol{A}^+)^3\right) \ \left(\text{resp. } T^- = \frac{1}{6}\operatorname{tr}\left((\boldsymbol{A}^-)^3\right)\right).$$

Armed with Proposition 2, if the true connectivity parameters α^+, β^+, α^-, β^- are unknown, the weight parameter ξ in Problem (9) can be replaced with its consistent estimator $\hat{\xi}$.

2.4 Optimization Algorithm and Its Recovery Guarantee

We observe that Problem (9) is analogous to the principal component analysis problem. It is thus natural to apply a power iteration-type method to solve it, which is described in Algorithm 1. Specifically, Algorithm 1 consists of two stages. In the first stage (lines 5–8), it uses the PIs to approximate the leading eigenvector of $\boldsymbol{W}$. In the second stage (lines 9–12), it employs the GPIs to iteratively refine the approximate eigenvector after proper scaling. In line 11, $\boldsymbol{W}\boldsymbol{x}^{t-1}/|\boldsymbol{W}\boldsymbol{x}^{t-1}|$ essentially computes the projection of $\boldsymbol{W}\boldsymbol{x}^{t-1}$ onto the set $\{1, -1\}^n$. Algorithm 1 is a generalization of the iterative method in Wang et al. (2020) for community recovery over unsigned graphs. Indeed, by setting $\xi = 0$ in line 2, Algorithm 1 reduces to the algorithm in Wang et al. (2020).

Next, we present the main theorem of this paper,

which establishes the exact recovery guarantee of Algorithm 1.

Theorem 1. *Suppose that Assumption 1 holds and $\alpha^+, \beta^+, \alpha^-, \beta^-$ satisfy (5). Then, for sufficiently large n, it holds with probability at least $1 - n^{-\Omega(1)}$ that Algorithm 1 outputs $\boldsymbol{x}^*$ or $-\boldsymbol{x}^*$ in $T_1 = \mathcal{O}(\log n/\log\log n)$ PIs and $T_2 = \mathcal{O}(\log n/\log\log n)$ GPIs.*

Theorem 1 indicates that under the information-theoretic limit (5), Algorithm 1 is able to exactly recover the ground-truth $\boldsymbol{x}^*$ up to a sign with high probability. Equipped with the iteration complexity given in Theorem 1, we can further obtain the time complexity of Algorithm 1.

Corollary 1. *Suppose that Assumption 1 holds and $\alpha^+, \beta^+, \alpha^-, \beta^-$ satisfy (5). Then, for sufficiently large n, it holds with probability at least $1 - n^{-\Omega(1)}$ that Algorithm 1 outputs $\boldsymbol{x}^*$ or $-\boldsymbol{x}^*$ in $\mathcal{O}(n\log^2 n/\log\log n)$ time.*

Corollary 1 indicates that Algorithm 1 achieves exact recovery in nearly linear time.

It should be pointed out that Algorithm 1 implicitly requires knowledge of the connectivity parameters $\alpha^+, \beta^+, \alpha^-, \beta^-$. Nevertheless, if these parameters are unknown, we can replace ξ with its consistent estimator $\hat{\xi}$ given by (11). Proposition 2 guarantees that the true ξ can be accurately estimated for sufficiently large n, thus the results in Theorem 1 and Corollary 1 still remain valid.

3 PROOFS OF MAIN RESULTS

In this section, we provide the major ingredients in the proofs of the results in Section 2. This involves estimating the connectivity parameters in the SSBM in Section 3.1 and analyzing the behavior of the PIs and GPIs in Sections 3.2 and 3.3, respectively. The detailed proofs can be found in the supplementary material.

3.1 Estimation of the Connectivity Parameters

We first present the following lemma about the concentration properties of the numbers of edges and triangles in $\mathcal{G}^+$ and $\mathcal{G}^-$, respectively.

Lemma 1. *Consider the same setting as Proposition 2. Suppose that $n \ge \max\{\alpha^+, \beta^-\}$. Then, it holds*

with probability at least $1 - n^{-\Omega(1)}$ that

$$\left|N^+ - a\left(\alpha^+ + \beta^+\right)\right| \le \left(\sqrt{n} + \alpha^+/2\right)\log n,$$
$$\left|N^- - a\left(\alpha^- + \beta^-\right)\right| \le \left(\sqrt{n} + \alpha^-/2\right)\log n,$$
$$\left|T^+ - b\left((\alpha^+)^3 + 3\alpha^+(\beta^+)^2\right)\right| \le 8(\alpha^+)^{\frac{9}{4}}\left(\log n\right)^{\frac{11}{4}},$$
$$\left|T^- - b\left((\alpha^-)^3 + 3\alpha^-(\beta^-)^2\right)\right| \le 8(\beta^-)^{\frac{9}{4}}\left(\log n\right)^{\frac{11}{4}},$$

where

$$a := \frac{n\log n}{4}, \quad b := \frac{\log^3 n}{24}. \tag{12}$$

Based on Lemma 1, we can obtain the non-asymptotic guarantee for the estimation of connectivity parameters.

Lemma 2. *Consider the same setting as Proposition 2. Suppose that $n \ge \max\{\alpha^+, \beta^-\}$. Then, it holds with probability at least $1 - n^{-\Omega(1)}$ that*

$$\max\left\{|\alpha^+ - \hat{\alpha}^+|, |\beta^+ - \hat{\beta}^+|\right\} \le \kappa_1 \left(\log n\right)^{-\frac{1}{12}}, \tag{13}$$

$$\max\left\{|\alpha^- - \hat{\alpha}^-|, |\beta^- - \hat{\beta}^-|\right\} \le \kappa_2 \left(\log n\right)^{-\frac{1}{12}}, \tag{14}$$

where $\kappa_1, \kappa_2 > 0$ are constants.

Equipped with Lemma 2, we can show Proposition 2, whose proof is deferred to Section C in the supplementary material.

3.2 Analysis of the Power Iterations

In our technical developments in Sections 3.2 and 3.3, we use the true parameters $\alpha^+, \beta^+, \alpha^-, \beta^-$ for conciseness of exposition. As discussed in the last paragraph of Section 2.4, this does not influence our main results. In this subsection, we analyze the performance of the PIs. We first present the spectral property of the matrix $\boldsymbol{W}$.

Lemma 3. *Suppose that Assumption 1 holds. Let $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ be the eigenvalues of $\boldsymbol{W}$ and $\boldsymbol{u}_1$ be the eigenvector associated with eigenvalue λ_1. Then, for all sufficiently large n, it holds with probability at least $1 - n^{-\Omega(1)}$ that*

$$|\lambda_1| \ge \frac{c_0}{2}\log n - 3c_1\sqrt{\log n},$$
$$|\lambda_i| \le 3c_1\sqrt{\log n}, \quad for\ i = 2, \ldots, n,$$

and

$$\min_{\theta \in \{\pm 1\}}\left\|\boldsymbol{u}_1 - \theta\frac{\boldsymbol{x}^*}{\sqrt{n}}\right\| \le \frac{c_2}{\sqrt{\log n}}, \tag{15}$$

where $c_0 = |(\alpha^+ - \beta^+) - \xi(\alpha^- - \beta^-)|$ and $c_1, c_2 > 0$ are constants.

Lemma 3 guarantees that the magnitude of the leading eigenvalue of $\boldsymbol{W}$ is sufficiently larger than other eigenvalues, which can be used to show that the leading eigenvector $\boldsymbol{u}_1$ is sufficiently close to the scaled ground-truth $\boldsymbol{x}^*/\sqrt{n}$ up to a sign. Further, we show in the next proposition that the iterates $\boldsymbol{y}^t$ in the first stage of Algorithm 1 converge to $\boldsymbol{u}_1$ at a linear rate.

Proposition 3. *Suppose that Assumption 1 holds. Let $\boldsymbol{u}_1$ be the leading eigenvector of $\boldsymbol{W}$ and $\{\boldsymbol{y}^t\}_{t \ge 0}$ be the sequence generated in the first stage of Algorithm 1. Then, for sufficiently large n, it holds with probability at least $1 - n^{-\Omega(1)}$ that*

$$\min_{\theta \in \{\pm 1\}}\left\|\boldsymbol{y}^t - \theta\boldsymbol{u}_1\right\| \le n\left(\frac{6c_1}{c_0\sqrt{\log n} - 6c_1}\right)^t,$$

for all $t \ge 0$, where c_0 and c_1 are the constants in Lemma 3.

3.3 Analysis of the Generalized Power Iterations

In this subsection, we study the convergence behavior of the GPIs. In particular, we establish the contraction property and finite termination property of the GPIs. We begin with a key lemma that says that the magnitude of each entry of $\boldsymbol{W}\boldsymbol{x}^*$ is at least in the order of $\log n$.

Lemma 4. *Suppose that Assumption 1 holds and $\alpha^+, \beta^+, \alpha^-, \beta^-$ satisfy (5). Then, for sufficiently large n, there exists a constant $\gamma > 0$ such that the following inequality holds with probability at least $1 - n^{-\Omega(1)}$:*

$$\min\{x_i^*(\boldsymbol{W}\boldsymbol{x}^*)_i : i = 1, \ldots, n\} \ge \gamma\log n. \tag{16}$$

We remark that this lemma cannot be directly implied by Abbe et al. (2020, Lemma 8), since it concerns categorical random variables rather than Bernoulli random variables. Moreover, this lemma holds at the information-theoretic limit, which is critical to the optimal recovery performance of Algorithm 1 and could be of independent interest. Next, we present a lemma that shows the contraction property of the GPIs.

Proposition 4. *Suppose that Assumption 1 holds and $\alpha^+, \beta^+, \alpha^-, \beta^-$ satisfy (5). Let $\boldsymbol{x} \in \mathbb{R}^n$ be such that $\|\boldsymbol{x}\| = \sqrt{n}$. Then, for sufficiently large n, either of the following two statements holds with probability at least $1 - n^{-\Omega(1)}$:*
(i) If $\|\boldsymbol{x} - \boldsymbol{x}^*\| \le 2c_2\sqrt{n/\log n}$, then we have

$$\left\|\frac{\boldsymbol{W}\boldsymbol{x}}{|\boldsymbol{W}\boldsymbol{x}|} - \boldsymbol{x}^*\right\| \le \frac{2c_3}{\gamma\sqrt{\log n}}\|\boldsymbol{x} - \boldsymbol{x}^*\|.$$

(ii) If $\|\boldsymbol{x} + \boldsymbol{x}^*\| \le 2c_2\sqrt{n/\log n}$, then we have

$$\left\|\frac{\boldsymbol{W}\boldsymbol{x}}{|\boldsymbol{W}\boldsymbol{x}|} + \boldsymbol{x}^*\right\| \le \frac{2c_3}{\gamma\sqrt{\log n}}\|\boldsymbol{x} + \boldsymbol{x}^*\|,$$

Xiaolu Wang*, Peng Wang†, Anthony Man-Cho So*

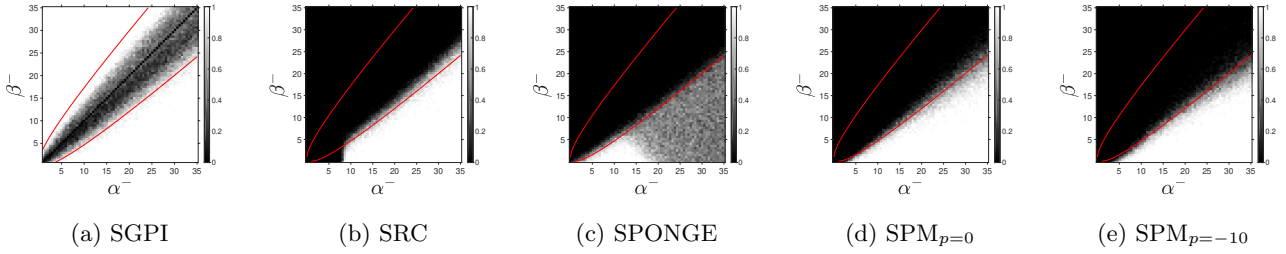(a) SGPI (b) SRC (c) SPONGE (d) SPM$_{p=0}$ (e) SPM$_{p=-10}$

Figure 1: Exact recovery ratios for different values of α^- and β^- with an increment of 0.5 (fixing $n = 300$, $\alpha^+ = 16$ and $\beta^+ = 9$). The information-theoretic threshold $(\sqrt{\alpha^-} - \sqrt{\beta^-})^2 = 1$ is plotted in red.
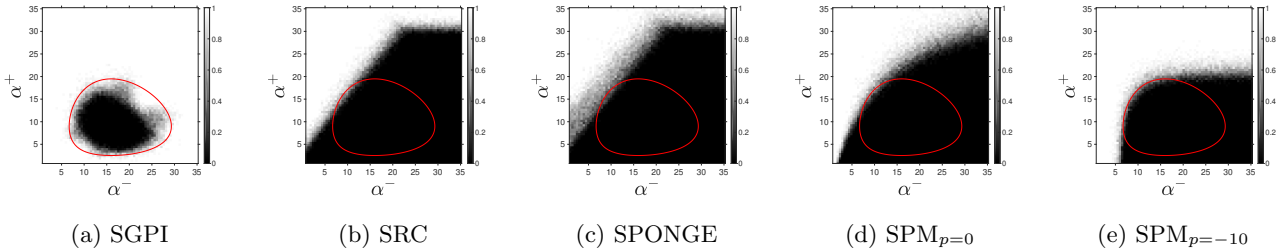

(a) SGPI (b) SRC (c) SPONGE (d) SPM$_{p=0}$ (e) SPM$_{p=-10}$

Figure 2: Exact recovery ratios for different values of α^+ and α^- with an increment of 0.5 (fixing $n = 300$, $\beta^+ = 9$ and $\beta^- = 16$). The information-theoretic threshold $(\sqrt{\alpha^+} - 3)^2 + (\sqrt{\alpha^-} - 4)^2 = 2$ is plotted in red.

where $c_3 = (6\sqrt{2}+1)c_1 + 2|\alpha^+ - \xi\alpha^-| + 1 + \xi$, c_1, c_2 are the constants in Lemma 3, and $\gamma > 0$ is the constant in Lemma 4.

Proposition 4 indicates that the iterates $\boldsymbol{x}^t$ in the second stage of Algorithm 1 converge linearly to either $\boldsymbol{x}^*$ or $-\boldsymbol{x}^*$ within the contraction region with radius $\mathcal{O}(\sqrt{n/\log n})$. Then, the following proposition characterizes a smaller one-step convergence region with radius 2, which further implies the finite termination of the GPIs.

Proposition 5. *Suppose that Assumption 1 holds and $\alpha^+, \beta^+, \alpha^-, \beta^-$ satisfy (5). Let $\boldsymbol{x} \in \{1, -1\}^n$. Then, for sufficiently large n, either of the following two statements holds with probability at least $1 - n^{-\Omega(1)}$:*
(i) If $\|\boldsymbol{x} - \boldsymbol{x}^*\| \leq 2$, then $\boldsymbol{W}\boldsymbol{x}/|\boldsymbol{W}\boldsymbol{x}| = \boldsymbol{x}^*$;
(ii) If $\|\boldsymbol{x} + \boldsymbol{x}^*\| \leq 2$, then $\boldsymbol{W}\boldsymbol{x}/|\boldsymbol{W}\boldsymbol{x}| = -\boldsymbol{x}^*$.

The results in Section 3.2 and 3.3 characterize the behavior of the iterates generated by Algorithm 1. In summary, the sequence $\{\boldsymbol{y}^t\}_{t\geq 0}$ in the first stage converges to $\boldsymbol{u}_1$ up to a sign (Proposition 3). Since $\boldsymbol{u}_1$ is close to $\boldsymbol{x}^*$ up to a sign (Lemma 3), the scaled iterate $\sqrt{n}\boldsymbol{y}^t$ will fall into the region of contraction around $\boldsymbol{x}^*$ or $-\boldsymbol{x}^*$ after sufficiently many PIs. Then, the PIs are switched to GPIs that will further approach $\boldsymbol{x}^*$ or $-\boldsymbol{x}^*$ (Proposition 4). The GPI terminates at $\boldsymbol{x}^*$ or $-\boldsymbol{x}^*$ once the iterate $\boldsymbol{x}^t$ goes into the one-step convergence region (Proposition 5). Consequently, we can analyze the overall iteration and time complexities of

Algorithm 1. The proofs of Theorem 1 and Corollary 1 are deferred to Section F in the supplementary material.

4 NUMERICAL EXPERIMENTS

In this section, we present the numerical results of our method. We validate the efficacy and efficiency of our Algorithm 1, which is referred to as SGPI in this section, by comparing it with three other state-of-the-art algorithms for signed graph clustering, which are Signed Ratio Cut (SRC) (Kunegis et al., 2010), Signed Positive Over Negative Generalized Eigenproblem (SPONGE) (Cucuringu et al., 2019), and Signed Power Means (SPM) (Mercado et al., 2019). Instead of using the ture connectivity parameters, we implement SGPI based on the estimated parameters given by (10a)–(10d) and (11). SPM contains the matrix power p as a hyperparameter. We take two typical values $p = 0$ and $p = -10$ for SPM, which are referred to as SPM$_{p=0}$ and SPM$_{p=-10}$, respectively. SPONGE has two hyperparameters τ^+ and τ^-, which we set as $\tau^+ = 10$ and $\tau^- = 1$. Our codes are implemented in MATLAB. For SPM, we directly use the MATLAB code provided by Mercado et al. (2019). All the experiments are conducted on a MacBook with 2.3GHz Intel Core i5 CPU and 8GB memory.

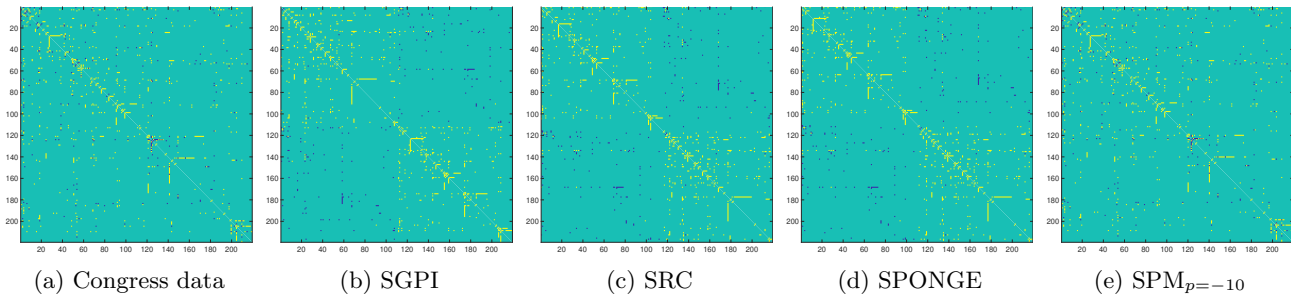| (a) Congress data | (b) SGPI | (c) SRC | (d) SPONGE | (e) SPM$_{p=-10}$ |

Figure 3: Original adjacency and the sorted adjacency matrices according to the communities identified by different algorithms (the yellow dots represent positive edges, the blue dots represent negative edges, and the green dots mean absence of edges)

n	100	500	1000	2000	5000
SGPI	**0.028**	**0.39**	**2.43**	**16.32**	**267.44**
SRC	0.034	0.54	3.07	23.40	312.95
SPONGE	0.151	1.70	8.87	66.68	832.80
SPM$_{p=0}$	0.562	3.69	17.05	113.34	1720.00
SPM$_{p=-10}$	0.386	2.07	9.91	65.11	964.41

Table 1: Total computation time (in seconds) of different community detection algorithms on synthetic data

	SGPI	SRC	SPONGE	SPM$_{p=-10}$
Congress	969	968	968	256
Highlandtribes	50.5	46.9	43.9	48.0

Table 2: Objective function values of algorithm outputs on real data

4.1 Phase Transition

We first report the community recovery performance under $\mathsf{SSBM}\left(n, \boldsymbol{x}^*, \frac{\alpha^+ \log n}{n}, \frac{\alpha^+ \log n}{n}, \frac{\alpha^+ \log n}{n}, \frac{\alpha^+ \log n}{n}\right)$ with $n = 300$. Given $\alpha^+, \alpha^-, \beta^+, \beta^-$, we independently generate 40 signed graphs for each algorithm and compute the ratio of exact recovery.

Figures 1 and 2 illustrate the heatmaps of exact recovery ratio of all tested algorithms. In Figure 1, all signed graphs are generated with fixed $\alpha^+ = 16$ and $\beta^+ = 9$. The information-theoretic limit (5) reduces to $(\sqrt{\alpha^-} - \sqrt{\beta^-})^2 \geq 1$. It can be observed that most algorithms exhibit obvious phase transition. Particularly, only our SGPI achieves exact recovery down to the information-theoretic limit, which supports the claim in Theorem 1. SRC and SPM performs well when $\sqrt{\beta^-} \leq \sqrt{\alpha^-} - 1$, while they can hardly recover the ground-truth if $\sqrt{\beta^-} \geq \sqrt{\alpha^-} + 1$. This is possibly because they equally treats the positive and negative edges of signed graphs, which cannot achieve exact recovery when there are even more negative edges within communities and less positive edges across communities. The superior recovery performance of our algorithm benefits from the weight parameter ξ in our formulation (9), which can be accurately estimated from the graphs.

In Figure 2, all signed graphs are generated with fixed $\beta^+ = 9$ and $\beta^- = 16$. The information-theoretic limit

(5) reduces to $(\sqrt{\alpha^+} - 3)^2 + (\sqrt{\alpha^-} - 4)^2 \geq 2$. We can observe analogous phase transition phenomenon in Figure 2. For the existing algorithms, SRC, SPONGE, and SPM$_{p=0}$ perform similarly, and SPM$_{p=-10}$ yields the best recovery performance. In contrast, our SGPI is the only algorithm that achieves exact recovery all the way down to the ring-shaped information-theoretic threshold.

4.2 Computational Efficiency

We compare the computation time of the aforementioned algorithms for community recovery over signed graphs with $\alpha^+ = 16$, $\alpha^- = 9$, $\beta^+ = 9$, $\beta^- = 16$. Given n, we independently generate 40 signed graphs for all algorithms and record their overall computation time for completing the community recovery. Specifically, the computation time of SGPI consists of the time consumed by parameter estimation, PIs, and GPIs. In Table 1, we report the total computation time consumed by each algorithm for different n. We can observe that SGPI is always the fastest algorithm to achieve community recovery over signed graphs of different sizes. In particular, SGPI exhibits comparable computation time performance with SRC and is substantially faster than other algorithms. Moreover, SGPI achieves the best scalability as n grows.

4.3 Experiments on Real Data

We conduct experiments to test the aforementioned algorithms on the Congress dataset (Thomas et al., 2006), which is a publicly available real-world signed

Xiaolu Wang[*], Peng Wang[†], Anthony Man-Cho So[*]

network in the KONECT project (Kunegis, 2013).[1] There are 219 nodes in the Congress dataset, whose adjacency matrix is presented in Figure 3a. Each node represents a politician and each edge represents a favorable (labeled by +1) or unfavorable (labeled by −1) mention of politicians speaking in the United States Congress. We apply the aforementioned algorithms to this signed network to identify the underlying community structure. The sorted adjacency matrices according to the communities identified by different algorithms are presented in Figures 3b to 3e. It can be observed that SGPI detects two obvious communities of similar sizes, where most positive edges appear within communities and most negative edges appear across communities. The sorted adjacency matrices of SRC and SPONGE exhibit similar community structures, while the sorted adjacency matrix of $SPM_{p=-10}$ exhibits less recognizable community structure.

We also evaluate the objective function of Problem (9) to assess the performance of the tested algorithms on real datasets. In addition to the Congress dataset, we also make evaluations on the Highlandtribes dataset (Read, 1954) in the KONECT project. The Highlandtribes network contains 16 tribes connected by friendship (corresponding to +1 edge) and enmity (corresponding to −1 edge). The results are reported in Table 2, where a higher function value usually indicates better performance. For the Congress dataset, the quantitative results coincide with the observations from the sorted adjacency matrices illustrated in Figure 3.

5   Conclusion

In this work, we proposed a new approach for community recovery over signed graphs based on the MLE formulation of the SSBM. We tackled the resulting non-convex and discrete optimization problem via a simple yet efficient two-stage iterative algorithm. Theoretical analysis shows that our method achieves exact recovery within nearly linear time at the information-theoretic limit. Numerical experiments validate the efficacy and efficiency of our approach and show its potential to identify communities in real networks. We leave it as a future work to consider community recovery in the signed SBMs with more than two communities.

Acknowledgements

[1]konect.cc/networks/convote/

References

E. Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(1):6446–6531, 2018.

E. Abbe and C. Sandon. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *Proceedings of 2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 670–688. IEEE, 2015.

E. Abbe, A. S. Bandeira, A. Bracher, and A. Singer. Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery. *IEEE Transactions on Network Science and Engineering*, 1(1):10–22, 2014.

E. Abbe, A. S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, 2015.

E. Abbe, J. Fan, K. Wang, and Y. Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *The Annals of Statistics*, 48(3):1452, 2020.

A. A. Amini and E. Levina. On semidefinite relaxations for the block model. *The Annals of Statistics*, 46(1):149–179, 2018.

N. Bansal, A. Blum, and S. Chawla. Correlation clustering. *Machine Learning*, 56(1):89–113, 2004.

F. Bonchi, E. Galimberti, A. Gionis, B. Ordozgoiti, and G. Ruffo. Discovering polarized communities in signed networks. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 961–970, 2019.

J. Bosch, P. Mercado, and M. Stoll. Node classification for signed networks using diffuse interface methods. *arXiv preprint arXiv:1809.06432*, 2018.

D. Cartwright and F. Harary. Structural balance: A generalization of Heider's theory. *Psychological Review*, 63(5):277, 1956.

K.-Y. Chiang, N. Natarajan, A. Tewari, and I. S. Dhillon. Exploiting longer cycles for link prediction in signed networks. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, pages 1157–1162, 2011.

K.-Y. Chiang, J. J. Whang, and I. S. Dhillon. Scalable clustering of signed networks using balance normalized cut. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pages 615–624, 2012.

K.-Y. Chiang, C.-J. Hsieh, N. Natarajan, I. S. Dhillon, and A. Tewari. Prediction and clustering in signed networks: A local to global perspective. *Journal of Machine Learning Research*, 15(1):1177–1213, 2014.

L. Chu, Z. Wang, J. Pei, J. Wang, Z. Zhao, and E. Chen. Finding gangs in war from signed networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1505–1514, 2016.

T. Coleman, J. Saunderson, and A. Wirth. A local-search 2-approximation for 2-correlation-clustering. In *Proceedings of the 16th Annual European Symposium on Algorithms*, pages 308–319. Springer, 2008.

M. Cucuringu, P. Davies, A. Glielmo, and H. Tyagi. SPONGE: A generalized eigenproblem for clustering signed networks. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, pages 1088–1098. PMLR, 2019.

J. A. Davis. Clustering and structural balance in graphs. *Human Relations*, 20(2):181–187, 1967.

T. Derr, Y. Ma, and J. Tang. Signed graph convolutional networks. In *Proceedings of the 2018 IEEE International Conference on Data Mining*, pages 929–934. IEEE, 2018.

T. Dittrich and G. Matz. Signal processing on signed graphs: Fundamentals and potentials. *IEEE Signal Processing Magazine*, 37(6):86–98, 2020.

A. Fujita, P. Severino, K. Kojima, J. R. Sato, A. G. Patriota, and S. Miyano. Functional clustering of time series gene expression data by granger causality. *BMC Systems Biology*, 6(1):1–12, 2012.

C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou. Achieving optimal misclassification proportion in stochastic block models. *Journal of Machine Learning Research*, 18(1):1980–2024, 2017.

I. Giotis and V. Guruswami. Correlation clustering with a fixed number of clusters. *Theory of Computing*, 2:249–266, 2006.

B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, 2016.

F. Harary. On the notion of balance of a signed graph. *Michigan Mathematical Journal*, 2(2):143–146, 1953.

V. Jog and P.-L. Loh. Information-theoretic bounds for exact recovery in weighted stochastic block models using the Renyi divergence. *arXiv preprint arXiv:1509.06418*, 2015.

J. Kunegis. KONECT: The koblenz network collection. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 1343–1350, 2013.

J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. W. De Luca, and S. Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In *Proceedings of the 2010 SIAM International Conference on Data Mining*, pages 559–570. SIAM, 2010.

G. Le Falher, N. Cesa-Bianchi, C. Gentile, and F. Vitale. On the troll-trust model for edge sign prediction in social networks. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pages 402–411. PMLR, 2017.

J. Lei and A. Rinaldo. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015.

P. Mercado, F. Tudisco, and M. Hein. Clustering signed networks with the geometric mean of Laplacians. In *Advances in Neural Information Processing Systems 29*, 2016.

P. Mercado, F. Tudisco, and M. Hein. Spectral clustering of signed graphs via matrix power means. In *Proceedings of the 36th International Conference on Machine Learning*, pages 4526–4536. PMLR, 2019.

E. Mossel, J. Neeman, and A. Sly. Consistency thresholds for binary symmetric block models. *arXiv preprint arXiv:1407.1591*, 2014.

G. J. Puleo and O. Milenkovic. Correlation clustering with constrained cluster sizes and extended weights bounds. *SIAM Journal on Optimization*, 25(3):1857–1872, 2015.

A. A. Quadeer, D. Morales-Jimenez, and M. R. McKay. Co-evolution networks of HIV/HCV are modular with direct association to structure and function. *PLoS Computational Biology*, 14(9):e1006409, 2018.

K. E. Read. Cultures of the Central Highlands, New Guinea. *Southwestern Journal of Anthropology*, 10(1):1–43, 1954.

M. Shahriari and M. Jalili. Ranking nodes in signed social networks. *Social Network Analysis and Mining*, 4(1):172, 2014.

M. Thomas, B. Pang, and L. Lee. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 327–335, 2006.

R.-C. Tzeng, B. Ordozgoiti, and A. Gionis. Discovering conflicting groups in signed networks. In *Advances in Neural Information Processing Systems 33*, 2020.

R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*, volume 47. Cambridge University Press, 2018.

V. H. Vu. Concentration of non-Lipschitz functions and applications. *Random Structures & Algorithms*, 20(3):262–316, 2002.

P. Wang, Z. Zhou, and A. M.-C. So. A nearly-linear time algorithm for exact community recovery in stochastic block model. In *Proceedings of the 37th International Conference on Machine Learning*, pages 10126–10135. PMLR, 2020.

P. Wang, H. Liu, Z. Zhou, and A. M.-C. So. Optimal non-convex exact recovery in stochastic block model via projected power method. In *In Proceedings of the 38th International Conference on Machine Learning*, pages 10828–10838. PMLR, 2021a.

P. Wang, Z. Zhou, and A. M.-C. So. Non-convex exact community recovery in stochastic block model. *Mathematical Programming*, pages 1–37, 2021b.

S. Wang, J. Tang, C. Aggarwal, Y. Chang, and H. Liu. Signed network embedding in social media. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, pages 327–335. SIAM, 2017.

S. Yardi and D. Boyd. Dynamic debates: An analysis of group polarization over time on Twitter. *Bulletin of Science, Technology & Society*, 30(5):316–327, 2010.

J. Ye, H. Cheng, Z. Zhu, and M. Chen. Predicting positive and negative links in signed social networks by transfer learning. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 1477–1488, 2013.

S. Yuan, X. Wu, and Y. Xiang. SNE: Signed network embedding. In *Proceedings of the 2017 Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 183–195. Springer, 2017.

S.-Y. Yun and A. Proutiere. Optimal cluster recovery in the labeled stochastic block model. In *Advances in Neural Information Processing Systems 29*, 2016.

Supplementary Material:
Exact Community Recovery over Signed Graphs

In the supplementary material, we the provide proofs of the technical results presented in Sections 2 and 3. We first introduce the notation that will be used.

Given an observed graph $\mathcal{G}$ generated according to $\mathsf{SSBM}(n, \boldsymbol{x}^*, p^+, p^-, q^+, q^-)$, we use N^+ (resp. N^-) to denote the number of positive (resp. negative) edges and T^+ (resp. T^-) to denote the number of positive (negative) triangles in $\mathcal{G}$. In addition, we use N_{in}^+ (resp. N_{in}^-) to denote the number of positive edges (resp. negative edges) within communities and N_{out}^+ (resp. N_{out}^-) to denote positive edges (resp. negative edges) across communities in $\mathcal{G}$.

A Proof of Proposition 1

Proof. Firstly, we note that

$$N^+ = N_{\text{in}}^+ + N_{\text{out}}^+, \quad N^- = N_{\text{in}}^- + N_{\text{out}}^-. \tag{17}$$

Let $\boldsymbol{x} \in \{1, -1\}^n$ be a vector whose i-th element indicates the community membership of node i. For equal-sized community detection, an admissible membership vector should also satisfy $\mathbf{1}^\top \boldsymbol{x} = 0$. Then, the likelihood function of $\boldsymbol{x}$, i.e., the probability of generating the graph $\mathcal{G}$ based on the community assignment $\boldsymbol{x}$, is

$$\mathbb{P}(\mathcal{G}|\boldsymbol{x}) \propto (p^+)^{N_{\text{in}}^+} (p^-)^{N_{\text{in}}^-} (1 - p^+ - p^-)^{\frac{n^2}{4} - \frac{n}{2} - N_{\text{in}}^+ - N_{\text{in}}^-} (q^+)^{N_{\text{out}}^+} (q^-)^{N_{\text{out}}^-} (1 - q^+ - q^-)^{\frac{n^2}{4} - N_{\text{out}}^+ - N_{\text{out}}^-}.$$

$$\propto \left(\frac{p^+}{q^+} \frac{1 - q^+ - q^-}{1 - p^+ - p^-} \right)^{N_{\text{in}}^+} \left(\frac{q^-}{p^-} \frac{1 - p^+ - p^-}{1 - q^+ - q^-} \right)^{N_{\text{out}}^-}.$$

Let

$$\mu_n = \log \left(\frac{p^+}{q^+} \frac{1 - q^+ - q^-}{1 - p^+ - p^-} \right), \ \nu_n = \log \left(\frac{q^-}{p^-} \frac{1 - p^+ - p^-}{1 - q^+ - q^-} \right),$$

then, it follows from (3) and (4) that

$$\mu_n = \log \left(\frac{\alpha^+}{\beta^+} \right) + \log \left(\frac{n - (\beta^+ + \beta^-) \log n}{n - (\alpha^+ + \alpha^-) \log n} \right), \ \nu_n = \log \left(\frac{\beta^-}{\alpha^-} \right) + \log \left(\frac{n - (\alpha^+ + \alpha^-) \log n}{n - (\beta^+ + \beta^-) \log n} \right).$$

Hence, we have

$$\log \mathbb{P}(\mathcal{G}|\boldsymbol{x}) \propto \left(\mu_n N_{\text{in}}^+ + \nu_n N_{\text{out}}^- \right). \tag{18}$$

According to (17), we note that

$$\boldsymbol{x}^\top \boldsymbol{A}^+ \boldsymbol{x} = \sum_{i,j \in [n]} A_{ij}^+ x_i x_j = \left(\sum_{i,j \in [n]: x_i x_j = 1} A_{ij}^+ \right) - \left(\sum_{i,j \in [n]: x_i x_j = -1} A_{ij}^+ \right) = 2N_{\text{in}}^+ - 2N_{\text{out}}^+ = 4N_{\text{in}}^+ - 2N^+$$

and

$$\boldsymbol{x}^\top \boldsymbol{A}^- \boldsymbol{x} = \sum_{i,j \in [n]} A_{ij}^- x_i x_j = \left(\sum_{i,j \in [n]: x_i x_j = 1} A_{ij}^- \right) - \left(\sum_{i,j \in [n]: x_i x_j = -1} A_{ij}^- \right) = 2N_{\text{in}}^- - 2N_{\text{out}}^- = 2N^- - 4N_{\text{out}}^-.$$

This, together with (18) and the fact that N^+, N^- are observed, yields the MLE formulation

$$\max \left\{ \boldsymbol{x}^\top (\boldsymbol{A}^+ - \xi_n \boldsymbol{A}^-) \boldsymbol{x} : \ \boldsymbol{x} \in \{1, -1\}^n, \ \mathbf{1}^\top \boldsymbol{x} = 0 \right\},$$

where $\xi_n = \nu_n / \mu_n$. □

B   Proofs in Section 3.1

B.1   Proof of Lemma 1

Proof. The numbers of edges in $\mathcal{G}^+$ and $\mathcal{G}^-$ satisfy

$$N^+ = \sum_{1 \leq i < j \leq n} A_{ij}^+, \quad N^- = \sum_{1 \leq i < j \leq n} A_{ij}^-, \tag{19}$$

respectively, where all the A_{ij}^+ in S_1^+ are i.i.d. $\mathbf{Bern}(p^+)$, all the A_{ij}^+ in S_2^+ are i.i.d. $\mathbf{Bern}(q^+)$ and independent of those in S_1^+, all the A_{ij}^- in S_1^- are i.i.d. $\mathbf{Bern}(p^-)$, and all the A_{ij}^- in S_2^- are i.i.d. $\mathbf{Bern}(q^-)$ and independent of those in S_1^-. Then, according to (3), (4), and (12), we have

$$\mathbb{E}[N^+] = \left(\frac{n^2}{4} - \frac{n}{2} \right) p^+ + \frac{n^2}{4} q^+ = a(\alpha^+ + \beta^+) - \frac{\alpha^+}{2} \log n, \tag{20}$$

$$\mathrm{Var}(N^+) = \left(\frac{n^2}{4} - \frac{n}{2} \right) p^+ (1 - p^+) + \frac{n^2}{4} q^+ (1 - q^+) \leq \frac{n^2}{2} p^+ = \frac{\alpha^+}{2} n \log n. \tag{21}$$

Then, by applying Bernstein's inequality for bounded distributions (see, e.g., Vershynin (2018, Theorem 2.8.4)) to (19), we obtain

$$\mathbb{P}\left(|N^+ - \mathbb{E}[N^+]| \geq \sqrt{n} \log n \right) \leq 2 \exp \left(- \frac{n \log^2 n / 2}{\alpha^+ n \log n / 2 + \sqrt{n} \log n / 3} \right)$$

$$\leq 2 \exp \left(- \frac{\log n}{\alpha^+ + 1/\sqrt{n}} \right)$$

$$= 2n^{- \frac{1}{\alpha^+ + 1/\sqrt{n}}}.$$

This implies that it holds with probability at least $1 - 2n^{- \frac{1}{\alpha^+ + 1/\sqrt{n}}}$ that

$$\left| N^+ - a\left(\alpha^+ + \beta^+ \right) + \frac{\alpha^+}{2} \log n \right| \leq \sqrt{n} \log n.$$

This, together with the triangle inequality, implies

$$\left| N^+ - a\left(\alpha^+ + \beta^+ \right) \right| \leq \sqrt{n} \log n + \frac{\alpha^+}{2} \log n.$$

Following similar arguments, we can show that it holds with probability at least $1 - 2n^{- \frac{1}{\beta^- + 1/\sqrt{n}}}$ that

$$\left| N^- - \left(a\alpha^- + b\beta^- \right) \right| \leq \sqrt{n} \log n + \frac{\alpha^-}{2} \log n.$$

Moreover, the numbers of triangles in $\mathcal{G}^+$ and $\mathcal{G}^-$, satisfy

$$T^+ = \frac{1}{6} \mathrm{tr} \left((\boldsymbol{A}^+)^3 \right), \ T_N^- = \frac{1}{6} \mathrm{tr} \left((\boldsymbol{A}^-)^3 \right). \tag{22}$$

respectively. Then, according to (3), (4), and (12), one can verify that

$$\mathbb{E}[T^+] = \frac{n}{6} \left(\frac{n}{2} - 1 \right) \left(\frac{n}{2} - 2 \right) \left(p^+ \right)^3 + \frac{n^2}{4} \left(\frac{n}{2} - 1 \right) p^+ \left(q^+ \right)^2$$

$$= \frac{\log^3 n}{24} \left(1 - \frac{6}{n} + \frac{8}{n^2} \right) (\alpha^+)^3 + \frac{\log^3 n}{8} \left(1 - \frac{2}{n} \right) \alpha^+ (\beta^+)^2$$

$$= b \left((\alpha^+)^3 + 3\alpha^+ (\beta^+)^2 \right) - \frac{\log^3 n}{2n} \left((\alpha^+)^3 + \alpha^+ (\beta^+)^2 \right) + \frac{\log^3 n}{3n^2} (\alpha^+)^3 \tag{23}$$

$$\leq \frac{(\alpha^+)^3}{6} \log^3 n. \tag{24}$$

where the inequality follows from $\alpha^+ > \beta^+$. This, together with Vu (2002, Theorem 4.2) with $k = 3$, $\tilde{k} = 2$, $\mathscr{E}_0 = \mathbb{E}[T^+]$, $\mathscr{E}_1 = \sqrt{\mathbb{E}[T^+]}$, $\mathscr{E}_2 = 1$, $\lambda = \log n$, $c_3 = 8\sqrt{6} + 2\sqrt{3} < 24$, and $d_3 = 14$, yields

$$\mathbb{P}\left(\left|T^+ - \mathbb{E}[T^+]\right| \geq 24\sqrt{\log n}\left(\mathbb{E}[T^+]\right)^{3/4}\right) \leq 14\exp\left(-\frac{\log n}{4}\right) = 14n^{-\frac{1}{4}}.$$

That is, it holds with probability at least $1 - 14n^{-\frac{1}{4}}$ that

$$\left|T^+ - \mathbb{E}[T^+]\right| \leq 24\sqrt{\log n}\left(\mathbb{E}[T^+]\right)^{3/4} \leq 7(\alpha^+)^{\frac{9}{4}}(\log n)^{\frac{11}{4}},$$

where the last inequality is due to (24). This, together with (23) and the triangle inequality, gives

$$\left|T^+ - b\left((\alpha^+)^3 + 3\alpha^+(\beta^+)^2\right)\right| \leq \left|T^+ - \mathbb{E}[T^+]\right| + \frac{\log^3 n}{2n}\left((\alpha^+)^3 + \alpha^+(\beta^+)^2\right) - \frac{\log^3 n}{3n^2}(\alpha^+)^3$$

$$\leq 7(\alpha^+)^{\frac{9}{4}}(\log n)^{\frac{11}{4}} + \frac{\alpha^{+3}\log^3 n}{2n} \leq 8(\alpha^+)^{\frac{9}{4}}(\log n)^{\frac{11}{4}},$$

where the last inequality holds for $n \geq \alpha^+$. Following similar arguments, when $n \geq \beta^-$, it holds with probability at least $1 - 14n^{-\frac{1}{4}}$ that

$$\left|T^- - b\left((\alpha^-)^3 + 3\alpha^-(\beta^-)^2\right)\right| \leq 8(\beta^-)^{\frac{9}{4}}(\log n)^{\frac{11}{4}}.$$

Finally, by applying the union bound, the desired results holds with probability at least $1 - 2n^{-\frac{1}{\alpha^+ + 1/\sqrt{n}}} - 2n^{-\frac{1}{\beta^- + 1/\sqrt{n}}} - 28n^{-\frac{1}{4}}$. $\qquad\square$

B.2 Proof of Lemma 2

Proof. Indeed, motivated by (20) and (23), we let the estimators of the connectivity parameters α^+ and β^+ be the solutions to the following system of equations:

$$\begin{cases} a\hat{\alpha}^+ + a\hat{\beta}^+ = N^+, \\ b(\hat{\alpha}^+)^3 + 3b\hat{\alpha}^+(\hat{\beta}^+)^2 = T^+. \end{cases} \tag{25}$$

Similarly, we let the estimators of the connectivity parameters α^- and β^- be the solutions to the following system of equations:

$$\begin{cases} a\hat{\alpha}^- + a\hat{\beta}^- = N^-, \\ b(\hat{\alpha}^-)^3 + 3b\hat{\alpha}^-(\hat{\beta}^-)^2 = T^-. \end{cases} \tag{26}$$

We can consider the following unified systems of equations for given $C, D \in \mathbb{R}$:

$$\begin{cases} ax + ay = C, \\ bx^3 + 3bxy^2 = D, \end{cases} \tag{27}$$

Eliminating y in the above system yields

$$4bx^3 - \frac{6bC}{a}x^2 + \frac{3bC^2}{a^2}x - D = 0.$$

According to (12), we further have

$$\left(\log^3 n\right)x^3 - \frac{6C\log^2 n}{n}x^2 + \frac{12C^2\log n}{n^2}x - 6D = 0.$$

This is equivalent to

$$\left(x\log n - \frac{2C}{n}\right)^3 = 6D - \frac{8C^3}{n^3}.$$

Xiaolu Wang*, **Peng Wang**[†], **Anthony Man-Cho So***

This implies that the solution of the system (27) has unique real root:

$$x = \frac{1}{\log n} \left(\frac{2C}{n} + \sqrt[3]{6D - \frac{8C^3}{n^3}} \right), \ y = \frac{C}{a} - x. \tag{28}$$

To proceed, let

$$\bar{N}^+ = a(\alpha^+ + \beta^+), \ \bar{T}^+ = b\alpha^{+^3} + 3b\alpha^+\beta^{+^2}, \ \delta_1 = N^+ - \bar{N}^+, \ \delta_2 = T^+ - \bar{T}^+.$$

By letting $C = \bar{N}^+$, $D = \bar{T}^+$ in (27), it is obvious that the real root is $x = \alpha^+, y = \beta^+$ and

$$\alpha^+ = \frac{1}{\log n} \left(\frac{2\bar{N}^+}{n} + \sqrt[3]{6\bar{T}^+ - \frac{8\bar{N}^{+^3}}{n^3}} \right). \tag{29}$$

In particular, it follows from (12) that

$$\sqrt[3]{6\bar{T}^+ - \frac{8\bar{N}^{+^3}}{n^3}} = \sqrt[3]{6\left(b\alpha^{+^3} + 3b\alpha^+\beta^{+^2}\right) - \frac{8\left(a(\alpha^+ + \beta^+)\right)^3}{n^3}} = \frac{\alpha^+ - \beta^+}{2} \log n \geq 0. \tag{30}$$

Moreover, by letting $C = N^+$ and $D = T^+$ in (27), it follows from (28) that

$$\hat{\alpha}^+ = \frac{1}{\log n} \left(\frac{2N^+}{n} + \sqrt[3]{6T^+ - \frac{8N^{+^3}}{n^3}} \right). \tag{31}$$

By Lemma 1, we have $|\delta_1| \lesssim \sqrt{n} \log n$ and $|\delta_2| \lesssim (\log n)^{\frac{11}{4}}$. This, together with $\bar{N}^+ \asymp n \log n$ and $\bar{T}^+ \asymp \log^3 n$, implies that for sufficiently large n,

$$\sqrt[3]{6T^+ - \frac{8N^{+^3}}{n^3}} = \sqrt[3]{6(\bar{T}^+ + \delta_2) - \frac{8(\bar{N}^+ + \delta_1)^3}{n^3}} \geq 0 \tag{32}$$

Then, we have

$$\left| \sqrt[3]{6\bar{T}^+ - \frac{8\bar{N}^{+^3}}{n^3}} - \sqrt[3]{6T^+ - \frac{8N^{+^3}}{n^3}} \right| \leq \left| \sqrt[3]{6(\bar{T}^+ - T^+) + \left(\frac{8N^{+^3}}{n^3} - \frac{8\bar{N}^{+^3}}{n^3} \right)} \right|$$

$$\leq \sqrt[3]{6\delta_2 + \frac{8\delta_1 \left((\bar{N}^+ + \delta_1)^2 + (\bar{N}^+ + \delta_1)\bar{N}^+ + \bar{N}^{+^2} \right)}{n^3}}$$

$$\leq \sqrt[3]{6|\delta_2|} + \frac{2}{n} \sqrt[3]{|\delta_1|(2\bar{N}^+ + \delta_1)^2}$$

$$\leq 4(\alpha^+)^{3/4}(\log n)^{11/12} + 4\sqrt[3]{\alpha^+ + \beta^+} \frac{\log n}{\sqrt[6]{n}},$$

where the first inequality is due to (30), (32), and the fact that $|\sqrt[3]{u} - \sqrt[3]{v}| \leq |\sqrt[3]{u - v}|$ for any $u, v \geq 0$ and the last inequality follows from Lemma 1. This, together with (29) and (31), gives

$$|\alpha^+ - \hat{\alpha}^+| = \frac{1}{\log n} \left| \frac{2(\bar{N}^+ - N^+)}{n} + \sqrt[3]{6\bar{T}^+ - \frac{8\bar{N}^{+^3}}{n^3}} - \sqrt[3]{6T^+ - \frac{8N^{+^3}}{n^3}} \right|$$

$$\leq 3n^{-1/2} + 4(\alpha^+)^{3/4}(\log n)^{-1/12} + 4\sqrt[3]{\alpha^+ + \beta^+} n^{-1/6}.$$

for sufficiently large n. Moreover, we have

$$\left| \beta^+ - \hat{\beta}^+ \right| = \left| \frac{\bar{N}^+ - N^+}{a} - (\alpha^+ - \hat{\alpha}^+) \right| \leq 6n^{-1/2} + 4(\alpha^+)^{3/4}(\log n)^{-1/12} + 4\sqrt[3]{\alpha^+ + \beta^+} n^{-1/6}.$$

Thus, there exists $\kappa_1 > 0$, such that for sufficiently large n,

$$\max\left\{ |\alpha^+ - \hat{\alpha}^+|, |\beta^+ - \hat{\beta}^+| \right\} \leq \kappa_1 (\log n)^{-\frac{1}{12}}.$$

Following similar arguments, we can show that there exists $\kappa_2 > 0$, such that for sufficiently large n,

$$\max\left\{ |\alpha^- - \hat{\alpha}^-|, |\beta^- - \hat{\beta}^-| \right\} \leq \kappa_2 (\log n)^{-\frac{1}{12}}.$$

$\square$

C   Proof of Proposition 2

Proof. By Lemma 2, for $n > \max\{\alpha^+, \beta^-\}$, it holds with probability at least $1 - n^{-\Omega(1)}$ that

$$|\alpha^+ - \hat{\alpha}^+| \leq \kappa_1 (\log n)^{-\frac{1}{12}},$$

which implies that

$$\log\left(1 - \frac{\kappa_1}{\alpha^+}(\log n)^{-\frac{1}{12}}\right) \leq \log\left(\frac{\hat{\alpha}^+}{\alpha^+}\right) \leq \log\left(1 + \frac{\kappa_1}{\alpha^+}(\log n)^{-\frac{1}{12}}\right),$$

for sufficiently large n such that $1 - \frac{\kappa_1}{\alpha^+}(\log n)^{-\frac{1}{12}} > 0$. Then, for sufficiently large n such that $1 - \frac{\kappa_1}{\alpha^+}(\log n)^{-\frac{1}{12}} > \frac{1}{2}$, we have

$$
\begin{aligned}
\zeta_1^+ := |\log \hat{\alpha}^+ - \log \alpha^+| &= \left|\log\left(\frac{\hat{\alpha}^+}{\alpha^+}\right)\right| \\
&\leq \max\left\{\left|\log\left(1 - \frac{\kappa_1}{\alpha^+}(\log n)^{-\frac{1}{12}}\right)\right|, \left|\log\left(1 + \frac{\kappa_1}{\alpha^+}(\log n)^{-\frac{1}{12}}\right)\right|\right\} \\
&\leq \frac{2\kappa_1}{\alpha^+}(\log n)^{-\frac{1}{12}},
\end{aligned}
\tag{33}
$$

where the second inequality is because $|\log(1-x)| = \log\left(1 + \frac{x}{1-x}\right) \leq \frac{x}{1-x} \leq 2x$ for $0 < x \leq \frac{1}{2}$ and $|\log(1+x)| \leq x$ for $x > 0$. Following similar arguments for (33), we can obtain

$$\zeta_2^+ := |\log \hat{\beta}^+ - \log \beta^+| \leq \frac{2\kappa_2}{\beta^+}(\log n)^{-\frac{1}{12}}, \tag{34}$$

$$\zeta_1^- := |\log \hat{\alpha}^- - \log \alpha^-| \leq \frac{2\kappa_1}{\alpha^-}(\log n)^{-\frac{1}{12}}, \tag{35}$$

$$\zeta_2^- := |\log \hat{\beta}^- - \log \beta^-| \leq \frac{2\kappa_2}{\beta^-}(\log n)^{-\frac{1}{12}}. \tag{36}$$

Therefore,

$$
\begin{aligned}
|\hat{\xi} - \xi| &= \left|\frac{\log(\hat{\beta}^-/\hat{\alpha}^-)}{\log(\hat{\alpha}^+/\hat{\beta}^+)} - \frac{\log(\beta^-/\alpha^-)}{\log(\alpha^+/\beta^+)}\right| \\
&= \left|\frac{\log\left(\frac{\alpha^+}{\beta^+}\right)\left(\log\left(\frac{\hat{\beta}^-}{\hat{\alpha}^-}\right) - \log\left(\frac{\beta^-}{\alpha^-}\right)\right) - \log\left(\frac{\beta^-}{\alpha^-}\right)\left(\log\left(\frac{\hat{\alpha}^+}{\hat{\beta}^+}\right) - \log\left(\frac{\alpha^+}{\beta^+}\right)\right)}{\log\left(\frac{\alpha^+}{\beta^+}\right)\log\left(\frac{\hat{\alpha}^+}{\hat{\beta}^+}\right)}\right| \\
&= \left|\frac{\log\left(\frac{\alpha^+}{\beta^+}\right)\left(\log \hat{\beta}^- - \log \beta^- + \log \alpha^- - \log \hat{\alpha}^-\right) - \log\left(\frac{\beta^-}{\alpha^-}\right)\left(\log \hat{\alpha}^+ - \log \alpha^+ + \log \beta^+ - \log \hat{\beta}^+\right)}{\log\left(\frac{\alpha^+}{\beta^+}\right)\left(\log \hat{\alpha}^+ - \log \hat{\beta}^+\right)}\right| \\
&\leq \left|\frac{\log\left(\frac{\alpha^+}{\beta^+}\right)}{\log\left(\frac{\alpha^+}{\beta^+}\right)\left(\log \hat{\alpha}^+ - \log \hat{\beta}^+\right)}\right|(\zeta_2^- + \zeta_1^-) + \left|\frac{\log\left(\frac{\beta^-}{\alpha^-}\right)}{\log\left(\frac{\alpha^+}{\beta^+}\right)\left(\log \hat{\alpha}^+ - \log \hat{\beta}^+\right)}\right|(\zeta_1^+ + \zeta_2^+) \\
&\leq \frac{1}{\log(\alpha^+/\beta^+)}\left(\frac{2\kappa_2}{\beta^-}(\log n)^{-\frac{1}{12}} + \frac{2\kappa_1}{\alpha^-}(\log n)^{-\frac{1}{12}}\right) + \frac{\xi}{\log(\alpha^+/\beta^+)}\left(\frac{2\kappa_1}{\alpha^+}(\log n)^{-\frac{1}{12}} + \frac{2\kappa_2}{\beta^+}(\log n)^{-\frac{1}{12}}\right) \\
&= \kappa(\log n)^{-\frac{1}{12}}
\end{aligned}
$$

where the first inequality follows from triangular inequality, the second inequality holds due to (33)–(36) and the fact that for sufficiently large n such that $\log\left(\frac{\alpha^+}{\beta^+}\right) - \left(\frac{2\kappa_1}{\alpha^+} + \frac{2\kappa_2}{\beta^+}\right)(\log n)^{-\frac{1}{12}} > 0$, we have

$$
\begin{aligned}
\left|\frac{\log\left(\frac{\alpha^+}{\beta^+}\right)}{\log\left(\frac{\alpha^+}{\beta^+}\right)\left(\log \hat{\alpha}^+ - \log \hat{\beta}^+\right)}\right| &= \left|\frac{1}{\log\left(\frac{\alpha^+}{\beta^+}\right) + (\log \hat{\alpha}^+ - \log \alpha^+) - (\log \hat{\beta}^+ - \log \beta^+)}\right| \\
&\leq \frac{1}{\log\left(\frac{\alpha^+}{\beta^+}\right) - \left(\frac{2\kappa_1}{\alpha^+} + \frac{2\kappa_2}{\beta^+}\right)(\log n)^{-\frac{1}{12}}} < \frac{1}{\log(\alpha^+/\beta^+)},
\end{aligned}
$$

and

$$\left| \frac{\log\left(\frac{\beta^-}{\alpha^-}\right)}{\log\left(\frac{\alpha^+}{\beta^+}\right)\left(\log\hat{\alpha}^+ - \log\hat{\beta}^+\right)} \right| = \left| \frac{\xi}{\log\left(\frac{\alpha^+}{\beta^+}\right) + (\log\hat{\alpha}^+ - \log\alpha^+) - (\log\hat{\beta}^+ - \log\beta^+)} \right|$$

$$\leq \frac{\xi}{\log\left(\frac{\alpha^+}{\beta^+}\right) - \left(\frac{2\kappa_1}{\alpha^+} + \frac{2\kappa_2}{\beta^+}\right)(\log n)^{-\frac{1}{12}}} < \frac{\xi}{\log(\alpha^+/\beta^+)},$$

and the last equality directly follows by setting

$$\kappa := \frac{1}{\log(\alpha^+/\beta^+)}\left(\frac{2\kappa_2}{\beta^-} + \frac{2\kappa_1}{\alpha^-}\right) + \frac{\xi}{\log(\alpha^+/\beta^+)}\left(\frac{2\kappa_1}{\alpha^+} + \frac{2\kappa_2}{\beta^+}\right).$$

The desired result is proved. □

D Proofs in Section 3.2

D.1 Required Technical Lemmas

We first present some technical lemmas that will be used in the sequel. The following lemma guarantees that the constant c_0 that appears throughout this paper is nonzero, which is necessary for many related results to remain valid.

Lemma 5. *If $\alpha^+ \neq \beta^+$, then $c_0 = (\alpha^+ - \beta^+) - \xi(\alpha^- - \beta^-) \neq 0$*

Proof. Suppose that $c_0 = 0$, then $\alpha^+ \neq \beta^+$ implies $\alpha^- \neq \beta^-$. By (8), we have

$$\frac{\alpha^+ - \beta^+}{\alpha^- - \beta^-} = \frac{\log\beta^- - \log\alpha^-}{\log\alpha^+ - \log\beta^+}. \tag{37}$$

1° If $\frac{\alpha^+ - \beta^+}{\alpha^- - \beta^-} > 0$, then we have

$$\alpha^+ - \beta^+ \text{ and } \alpha^- - \beta^- \text{ have the same sign}$$
$$\Leftrightarrow \log\alpha^+ - \log\beta^+ \text{ and } \log\alpha^- - \log\beta^- \text{ have the same sign}$$
$$\Leftrightarrow \frac{\log\alpha^+ - \log\beta^+}{\log\alpha^- - \log\beta^-} > 0 \Leftrightarrow \frac{\log\beta^- - \log\alpha^-}{\log\alpha^+ - \log\beta^+} < 0,$$

which contradicts (37).

2° If $\frac{\alpha^+ - \beta^+}{\alpha^- - \beta^-} < 0$, similar arguments yield a contradiction.

As a consequence, $c_0 \neq 0$. □

Lemma 6. *Suppose that Assumption 1 holds and $\rho = \mathbf{1}^\top \widetilde{A} \mathbf{1}/n^2$. Then, it holds with probability at least $1 - 2n^{-\frac{1}{2(\alpha^+ + \beta^+ + 1)}} - 2n^{-\frac{1}{2(\alpha^- + \beta^- + 1)}}$ that*

$$|\rho - \mathbb{E}[\rho]| \leq \frac{2\log n}{n^{3/2}}, \tag{38}$$

where $\mathbb{E}[\rho] = \frac{(p^+ + q^+) - \xi(p^- + q^-)}{2} - \frac{p^+ - \xi p^-}{n}$.

Proof. For ease of exposition, we define

$$\rho^+ = \frac{\mathbf{1}^\top A^+ \mathbf{1}}{n^2}, \quad \rho^- = \frac{\mathbf{1}^\top A^- \mathbf{1}}{n^2}.$$

Since Assumption 1 holds, then by (6), we can verify that

$$\mathbb{E}[\rho^+] = \frac{p^+ + q^+}{2} - \frac{p^+}{n}, \quad \mathbb{E}[\rho^-] = \frac{p^- + q^-}{2} - \frac{p^-}{n}.$$

Thus, according to Wang et al. (2020, Lemma 1), we obtain

$$\mathbb{P}\left(\left|\rho^+ - \mathbb{E}[\rho^+]\right| \leq \frac{\log n}{n^{3/2}}\right) = \mathbb{P}\left(\left|\rho^+ - \left(\frac{1}{2}(p^+ + q^+) - \frac{p^+}{n}\right)\right| \leq \frac{\log n}{n^{3/2}}\right) \geq 1 - 2n^{-\frac{1}{2(\alpha^+ + \beta^+ + 1)}}$$

and

$$\mathbb{P}\left(\left|\rho^- - \mathbb{E}[\rho^-]\right| \leq \frac{\log n}{n^{3/2}}\right) = \mathbb{P}\left(\left|\rho^- - \left(\frac{1}{2}(p^- + q^-) - \frac{p^-}{n}\right)\right| \leq \frac{\log n}{n^{3/2}}\right) \geq 1 - 2n^{-\frac{1}{2(\alpha^- + \beta^- + 1)}}.$$

These, together with $\rho = \rho^+ - \xi\rho^-$ and the union bound, imply that it holds with probability at least $1 - 2n^{-\frac{1}{2(\alpha^+ + \beta^+ + 1)}} - 2n^{-\frac{1}{2(\alpha^- + \beta^- + 1)}}$ that

$$|\rho - \mathbb{E}[\rho]| \leq \left|\rho^+ - \mathbb{E}[\rho^+]\right| + \xi\left|\rho^- - \mathbb{E}[\rho^-]\right| \leq \frac{(1 + \xi)\log n}{n^{3/2}}, \tag{39}$$

where

$$\mathbb{E}[\rho] = \frac{1}{2}\left(p^+ + q^+\right) - \frac{\xi}{2}\left(p^- + q^-\right) - \frac{1}{n}\left(p^+ - \xi p^-\right). \tag{40}$$

This complete the proof. □

Next, we present a spectral bound on the deviation of $\widetilde{\boldsymbol{A}}$ from its mean.

Lemma 7. *Suppose that Assumption 1 holds. Then, it holds with probability at least $1 - 2n^{-3}$ that*

$$\|\widetilde{\boldsymbol{A}} - \mathbb{E}[\widetilde{\boldsymbol{A}}]\| \leq c_1\sqrt{\log n}, \tag{41}$$

where $c_1 > 0$ is a constant.

Proof. According to Lei and Rinaldo (2015, Theorem 5.2) and the union bound, it holds with probability at least $1 - 2n^{-3}$ for some constants $c_1^+, c_1^- > 0$ that

$$\|\boldsymbol{A}^+ - \mathbb{E}[\boldsymbol{A}^+]\| \leq c_1^+\sqrt{\log n}, \quad \|\boldsymbol{A}^- - \mathbb{E}[\boldsymbol{A}^-]\| \leq c_1^-\sqrt{\log n}.$$

These, together with $\widetilde{\boldsymbol{A}} = \boldsymbol{A}^+ - \xi\boldsymbol{A}^-$, yield

$$\begin{aligned}
\|\widetilde{\boldsymbol{A}} - \mathbb{E}[\widetilde{\boldsymbol{A}}]\| &= \|\boldsymbol{A}^+ - \xi\boldsymbol{A}^- - \mathbb{E}[\boldsymbol{A}^+ - \xi\boldsymbol{A}^-]\| \\
&\leq \|\boldsymbol{A}^+ - \mathbb{E}[\boldsymbol{A}^+]\| + \xi\|\boldsymbol{A}^- - \mathbb{E}[\boldsymbol{A}^-]\| \\
&\leq c_1^+\sqrt{\log n} + \xi c_1^-\sqrt{\log n} \\
&= c_1\sqrt{\log n},
\end{aligned}$$

where the last equality holds by letting $c_1 = c_1^+ + \xi c_1^- = c_1^+ + \frac{\log(\beta^-/\alpha^-)}{\log(\alpha^+/\beta^+)}c_1^- > 0$. □

D.2 Proof of Lemma 3 (Spectral Gap)

Proof. Since Assumption 1 holds, then by (6), we have

$$\mathbb{E}[\boldsymbol{A}^+] = \frac{p^+ + q^+}{2}\boldsymbol{E} + \frac{p^+ - q^+}{2}\boldsymbol{x}^*\boldsymbol{x}^{*\top} - p^+\boldsymbol{I}, \quad \mathbb{E}[\boldsymbol{A}^-] = \frac{p^- + q^-}{2}\boldsymbol{E} + \frac{p^- - q^-}{2}\boldsymbol{x}^*\boldsymbol{x}^{*\top} - p^-\boldsymbol{I}.$$

These, together with $\boldsymbol{W} = \widetilde{\boldsymbol{A}} - \rho\boldsymbol{E}$, $\widetilde{\boldsymbol{A}} = \boldsymbol{A}^+ - \xi\boldsymbol{A}^-$, and (40), give

$$\begin{aligned}
\mathbb{E}[\boldsymbol{W}] &= \mathbb{E}[\widetilde{\boldsymbol{A}}] - \mathbb{E}[\rho]\boldsymbol{E} = \mathbb{E}[\boldsymbol{A}^+] - \xi\mathbb{E}[\boldsymbol{A}^-] - \mathbb{E}[\rho]\boldsymbol{E} \\
&= \frac{1}{2}\left((p^+ - q^+) - \xi(p^- - q^-)\right)\boldsymbol{x}^*\boldsymbol{x}^{*\top} + \frac{1}{n}(p^+ - \xi p^-)\boldsymbol{E} - (p^+ - \xi p^-)\boldsymbol{I}. \tag{42}
\end{aligned}$$

Xiaolu Wang[*], Peng Wang[†], Anthony Man-Cho So[*]

Let $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_n$ be the eigenvalues of $((p^+ - q^+) - \xi(p^- - q^-)) \, \boldsymbol{x}^* \boldsymbol{x}^{*\top}/2$. Using (3) and (4), one can compute

$$\nu_1 = \frac{\log n}{2} \left((\alpha^+ - \beta^+) - \xi(\alpha^- - \beta^-)\right), \quad \nu_i = 0 \text{ for } i = 2, \ldots, n. \tag{43}$$

Let $\boldsymbol{\Delta} := \boldsymbol{W} - ((p^+ - q^+) - \xi(p^- - q^-)) \, \boldsymbol{x}^* \boldsymbol{x}^{*\top}/2$, then it follows from Weyl's inequality that

$$||\lambda_i| - |\nu_i|| \leq |\lambda_i - \nu_i| \leq \|\boldsymbol{\Delta}\|, \ i = 1, 2, \cdots, n. \tag{44}$$

Besides, note that

$$\boldsymbol{W} = \mathbb{E}[\boldsymbol{W}] + (\widetilde{\boldsymbol{A}} - \mathbb{E}[\widetilde{\boldsymbol{A}}]) - (\rho - \mathbb{E}[\rho]) \, \boldsymbol{E}. \tag{45}$$

According to Lemma 6, Lemma 7, and the union bound, it holds with probability at least $1 - 2n^{-3} - 2n^{-\frac{1}{2(\alpha^+ + \beta^+ + 1)}} - 2n^{-\frac{1}{2(\alpha^- + \beta^- + 1)}}$

$$\|(\widetilde{\boldsymbol{A}} - \mathbb{E}[\widetilde{\boldsymbol{A}}]) - (\rho - \mathbb{E}[\rho]) \, \boldsymbol{E}\| \leq \|\widetilde{\boldsymbol{A}} - \mathbb{E}[\widetilde{\boldsymbol{A}}]\| + n \, |\rho - \mathbb{E}[\rho]| \leq c_1 \sqrt{\log n} + (1 + \xi) \frac{\log n}{\sqrt{n}} \leq 2c_1 \sqrt{\log n},$$

where the last inequality is due to $n \geq (1 + \xi)^2 \log n / c_1^2$ for sufficiently large n. This, together with (42) and (45), implies

$$\|\boldsymbol{\Delta}\| = \left\|\boldsymbol{W} - \frac{1}{2} \left((p^+ - q^+) - \xi(p^- - q^-)\right) \boldsymbol{x}^* \boldsymbol{x}^{*\top}\right\| \leq 2c_1 \sqrt{\log n} + 2|p^+ - \xi p^-| \leq 3c_1 \sqrt{\log n} \tag{46}$$

where the last inequality is due to (3) and (8) for sufficiently large n. By (43), (44), and (46), for all sufficiently large n, it holds with probability at least $1 - n^{-\Omega(1)}$ that

$$|\lambda_1| \geq |\nu_1| - \|\boldsymbol{\Delta}\| \geq \frac{c_0}{2} \log n - 3c_1 \sqrt{\log n}, \tag{47}$$

$$|\lambda_i| \leq |\nu_i| + \|\boldsymbol{\Delta}\| \leq 3c_1 \sqrt{\log n}, \text{ for } i = 2, \ldots, n. \tag{48}$$

Moreover, applying the Davis-Kahan theorem (see, e.g., Vershynin (2018, Theorem 4.5.5)) to $\boldsymbol{W}$ yields

$$\min_{\theta \in \{\pm 1\}} \left\|\boldsymbol{u}_1 - \theta \frac{\boldsymbol{x}^*}{\sqrt{n}}\right\| \leq \frac{2\sqrt{2} \left\|\boldsymbol{W} - \frac{1}{2}\left((p^+ - q^+) - \xi(p^- - q^-)\right) \boldsymbol{x}^* \boldsymbol{x}^{*\top}\right\|}{|\nu_1|} \leq \frac{12\sqrt{2} c_1}{|(\alpha^+ - \beta^+) - \xi(\alpha^- - \beta^-)| \sqrt{\log n}},$$

where the last inequality is due to (43) and (46), and the last inequality holds for sufficiently large n and for $i = 2, \ldots, n$. Setting $c_2 := 12\sqrt{2} c_1 / c_0$ gives (15). $\qquad\square$

D.3 Proof of Proposition 3

Proof. Let $\boldsymbol{W} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^\top$ be the eigenvalue decomposition of of $\boldsymbol{W}$, where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 \geq \cdots \geq \lambda_n$ being the eigenvalues and $\boldsymbol{U} = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{bmatrix}$ are the associated eigenvectors. Suppose that the results in Lemma 3 holds, which happens with probability at least $1 - n^{-\Omega(1)}$. First, by the power iteration in Algorithm 1, we have

$$\boldsymbol{y}^{(t)} = \frac{\boldsymbol{W} \boldsymbol{y}^{(t-1)}}{\|\boldsymbol{W} \boldsymbol{y}^{(t-1)}\|_2} = \frac{\boldsymbol{W}^2 \boldsymbol{y}^{(t-2)}}{\|\boldsymbol{W}^2 \boldsymbol{y}^{(t-2)}\|_2} = \cdots = \frac{\boldsymbol{W}^t \boldsymbol{y}^{(0)}}{\|\boldsymbol{W}^t \boldsymbol{y}^{(0)}\|_2}.$$

Let the eigendecomposition of $\boldsymbol{W}$ be $\boldsymbol{W} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^\top$ and $\boldsymbol{b} = \boldsymbol{U}^\top \boldsymbol{y}^{(0)}$, then

$$\langle \boldsymbol{u}_1, \boldsymbol{y}^{(t)} \rangle = \frac{\boldsymbol{u}_1^\top \boldsymbol{W}^t \boldsymbol{y}^{(0)}}{\|\boldsymbol{W}^t \boldsymbol{y}^{(0)}\|_2} = \frac{\boldsymbol{u}_1^\top \boldsymbol{U} \boldsymbol{\Lambda}^t \boldsymbol{U}^\top \boldsymbol{y}^{(0)}}{\|\boldsymbol{U} \boldsymbol{\Lambda}^t \boldsymbol{U}^\top \boldsymbol{y}^{(0)}\|_2} = \frac{\boldsymbol{u}_1^\top \boldsymbol{U} \boldsymbol{\Lambda}^t \boldsymbol{b}}{\|\boldsymbol{\Lambda}^t \boldsymbol{b}\|_2} = \frac{\lambda_1^t b_1}{\sqrt{\sum_{i=1}^n \lambda_i^{2t} b_i^2}}.$$

Besides, Cauchy-Schwartz inequality reads

$$|\langle \boldsymbol{u}_1, \boldsymbol{y}^{(t)} \rangle| \leq \|\boldsymbol{u}_1\|_2 \|\boldsymbol{y}^{(t)}\|_2 = 1.$$

We then consider the following two cases:

1° Assume that $b_1 \geq 0$. Since $\lambda_1 > 0$, then $\langle \boldsymbol{u}_1, \boldsymbol{y}^{(t)} \rangle \geq 0$, i.e., $0 \leq \langle \boldsymbol{u}_1, \boldsymbol{y}^{(t)} \rangle \leq 1$. By Wang et al. (2020, Lemma 4), for sufficiently large n, it holds with probability at least $1 - 2n^{-1/2}$ that

$$\sum_{i=2}^{n} \left(\frac{b_i}{b_1} \right)^2 \leq \frac{n^2}{2}. \tag{49}$$

Then, for sufficiently large n, it holds with probability at least $1 - n^{-\Omega(1)}$ that

$$\|\boldsymbol{y}^{(t)} - \boldsymbol{u}_1\|_2^2 = \|\boldsymbol{y}^{(t)}\|_2^2 + \|\boldsymbol{u}_1\|_2^2 - 2\langle \boldsymbol{u}_1, \boldsymbol{y}^{(t)} \rangle = 2 \left(1 - \langle \boldsymbol{u}_1, \boldsymbol{y}^{(t)} \rangle \right)$$

$$\leq 2 \left(1 - \langle \boldsymbol{u}_1, \boldsymbol{y}^{(t)} \rangle^2 \right) = 2 \frac{\sum_{i=2}^{n} \lambda_i^{2t} b_i^2}{\sum_{i=1}^{n} \lambda_i^{2t} b_i^2} \leq 2 \frac{\sum_{i=2}^{n} \lambda_i^{2t} b_i^2}{\lambda_1^{2k} b_1^2} \leq 2 \left| \frac{\lambda'}{\lambda_1} \right|^{2t} \sum_{i=2}^{n} \left(\frac{b_i}{b_1} \right)^2$$

$$\leq \left(\frac{6c_1}{|\alpha^+ - \beta^+ - \xi(\alpha^- - \beta^-)|\sqrt{\log n} - 6c_1} \right)^{2t} n^2,$$

where $\lambda' := \max\{|\lambda_2|, \ldots, |\lambda_n|\}$ and the last inequality follows from (47), (48), and (49). That is,

$$\|\boldsymbol{y}^{(t)} - \boldsymbol{u}_1\|_2 \leq n \left(\frac{6c_1}{|\alpha^+ - \beta^+ - \xi(\alpha^- - \beta^-)|\sqrt{\log n} - 6c_1} \right)^t.$$

2° Assume that $b_1 < 0$. Since $\lambda_1 > 0$, then $\langle \boldsymbol{u}_1, \boldsymbol{y}^{(t)} \rangle < 0$, i.e., $-1 \leq \langle \boldsymbol{u}_1, \boldsymbol{y}^{(t)} \rangle < 0$. Following similar arguments with case 1°, we have

$$\|\boldsymbol{y}^{(t)} + \boldsymbol{u}_1\|_2^2 \leq 2 \left(1 - \langle \boldsymbol{u}_1, \boldsymbol{y}^{(t)} \rangle^2 \right) \leq \left(\frac{6c_1}{|\alpha^+ - \beta^+ - \xi(\alpha^- - \beta^-)|\sqrt{\log n} - 6c_1} \right)^{2t} n^2,$$

i.e.,

$$\|\boldsymbol{y}^{(t)} + \boldsymbol{u}_1\|_2 \leq n \left(\frac{6c_1}{|\alpha^+ - \beta^+ - \xi(\alpha^- - \beta^-)|\sqrt{\log n} - 6c_1} \right)^t.$$

Combining cases 1° and 2°, we have for all $t \geq 0$ and for sufficiently large n, it holds with probability at least $1 - n^{-\Omega(1)}$ that

$$\min_{\theta \in \{\pm 1\}} \|\boldsymbol{y}^{(t)} - \theta \boldsymbol{u}_1\|_2 \leq n \left(\frac{6c_1}{c_0 \sqrt{\log n} - 6c_1} \right)^t.$$

$\square$

E   Proofs in Section 3.3

E.1   Proof of Lemma 4

Proof. Let $m := n/2$ and

$$I_1 := \{(i,j) : \text{nodes } i, j \text{ belong to the same community}\},$$
$$I_2 := \{(i,j) : \text{nodes } i, j \text{ belong to different communities}\}.$$

Then, it follows from $\boldsymbol{E}\boldsymbol{x}^* = \boldsymbol{0}$, (3), (4), and (6) that

$$x_i^* (\boldsymbol{W}\boldsymbol{x}^*)_i = x_i^* (\widetilde{\boldsymbol{A}}\boldsymbol{x}^*)_i = x_i^* \left((\boldsymbol{A}^+ - \xi \boldsymbol{A}^-)\boldsymbol{x}^* \right)_i$$

$$= \sum_{j:(i,j)\in I_1} (A_{ij}^+ - \xi A_{ij}^-) x_i^* x_j^* + \sum_{j:(i,j)\in I_2} (A_{ij}^+ - \xi A_{ij}^-) x_i^* x_j^* = \sum_{j=1}^{m-1} X_j - \sum_{j=1}^{m} Y_j,$$

Xiaolu Wang[*], **Peng Wang**[†], **Anthony Man-Cho So**[*]

where $\{X_j : j = 1, \ldots, m-1\}$ are i.i.d. random variables whose distributions are given by

$$X_j = \begin{cases} 1, & \text{w.p. } p^+, \\ -\xi, & \text{w.p. } p^-, \\ 0, & \text{w.p. } 1 - (p^+ + p^-), \end{cases} \quad \text{for } j = 1, \ldots, m, \tag{50}$$

and $\{Y_j : j = 1, \ldots, m\}$, independent of $\{X_j : j = 1, \ldots, m-1\}$, are i.i.d. random variables whose distributions are given by

$$Y_j = \begin{cases} 1, & \text{w.p. } q^+, \\ -\xi, & \text{w.p. } q^-, \\ 0, & \text{w.p. } 1 - (q^+ + q^-), \end{cases} \quad \text{for } j = 1, \ldots, m. \tag{51}$$

For any $\gamma > 0$, applying Markov's inequality to the moment generating function with $\lambda < 0$ yields

$$\mathbb{P}\left(x_i^*(\boldsymbol{W}\boldsymbol{x}^*)_i \leq \gamma \log n\right) = \mathbb{P}\left(\sum_{j=1}^{m-1} X_j - \sum_{j=1}^{m} Y_j \leq \gamma \log n\right) \leq \mathbb{P}\left(\sum_{j=1}^{m} X_j - \sum_{j=1}^{m} Y_j \leq \gamma \log n + 1\right)$$

$$\leq \mathbb{P}\left(\sum_{j=1}^{m} X_j - \sum_{j=1}^{m} Y_j \leq \gamma' \log n\right) = \mathbb{P}\left(\exp\left(\lambda\left(\sum_{j=1}^{m} X_j - \sum_{j=1}^{m} Y_j\right)\right) \geq \exp(\lambda\gamma' \log n)\right)$$

$$\leq n^{-\lambda\gamma'} \mathbb{E}\left[\exp\left(\lambda\left(\sum_{j=1}^{m} X_j - \sum_{j=1}^{m} Y_j\right)\right)\right]$$

$$= n^{-\lambda\gamma'} \mathbb{E}\left[\exp\left(\lambda\sum_{j=1}^{m} X_j\right)\right] \mathbb{E}\left[\exp\left(-\lambda\sum_{j=1}^{m} Y_j\right)\right]$$

$$= n^{-\lambda\gamma'} \prod_{j=1}^{m} \mathbb{E}\left[\exp(\lambda X_j)\right] \prod_{j=1}^{m} \mathbb{E}\left[\exp(-\lambda Y_j)\right]$$

$$= n^{-\lambda\gamma'} \left(p^+ e^\lambda + p^- e^{-\xi\lambda} + 1 - (p^+ + p^-)\right)^{\frac{n}{2}} \left(q^+ e^{-\lambda} + q^- e^{\xi\lambda} + 1 - (q^+ + q^-)\right)^{\frac{n}{2}},$$

where the second inequality holds by $\gamma' \geq \gamma + 1/\log n$, the third equality is due to the fact that $\{Y_j : j = 1, \ldots, m\}$ are independent of $\{X_j : j = 1, \ldots, m\}$, and the last equality follows from (50) and (51). Taking logarithm of the above inequality and using (3) and (4) yield

$$\log \mathbb{P}\left(x_i^*(\boldsymbol{W}\boldsymbol{x}^*)_i \leq \gamma \log n\right)$$

$$\leq -\lambda\gamma' \log n + \frac{n}{2} \log\left(\frac{\alpha^+ \log n}{n}(e^\lambda - 1) + \frac{\alpha^- \log n}{n}(e^{-\xi\lambda} - 1) + 1\right)$$

$$+ \frac{n}{2} \log\left(\frac{\beta^+ \log n}{n}(e^{-\lambda} - 1) + \frac{\beta^- \log n}{n}(e^{\xi\lambda} - 1) + 1\right)$$

$$\leq -\lambda\gamma' \log n + \frac{n}{2}\left(\frac{\alpha^+ \log n}{n}(e^\lambda - 1) + \frac{\alpha^- \log n}{n}(e^{-\xi\lambda} - 1)\right)$$

$$+ \frac{n}{2}\left(\frac{\beta^+ \log n}{n}(e^{-\lambda} - 1) + \frac{\beta^- \log n}{n}(e^{\xi\lambda} - 1)\right)$$

$$= \left(-\lambda\gamma' + \frac{\alpha^+}{2}(e^\lambda - 1) + \frac{\alpha^-}{2}(e^{-\xi\lambda} - 1) + \frac{\beta^+}{2}(e^{-\lambda} - 1) + \frac{\beta^-}{2}(e^{\xi\lambda} - 1)\right) \log n, \tag{52}$$

where the second inequality is due to $\log(1 + x) \leq x$ for any $x > -1$. Taking $\lambda = -\log(\alpha^+/\beta^+)/2 < 0$, we have

$$e^\lambda = \exp\left(-\frac{\log(\alpha^+/\beta^+)}{2}\right) = \sqrt{\frac{\beta^+}{\alpha^+}}, \quad e^{\xi\lambda} = \exp\left(-\frac{\log(\beta^-/\alpha^-)}{\log(\alpha^+/\beta^+)} \cdot \frac{\log(\alpha^+/\beta^+)}{2}\right) = \sqrt{\frac{\alpha^-}{\beta^-}}.$$

It follows from this that

$$\frac{\alpha^+}{2}(e^\lambda - 1) + \frac{\alpha^-}{2}(e^{\xi\lambda} - 1) + \frac{\beta^+}{2}(e^{-\lambda} - 1) + \frac{\beta^-}{2}(e^{-\xi\lambda} - 1)$$

$$= \frac{\alpha^+}{2}\left(\sqrt{\frac{\beta^+}{\alpha^+}} - 1\right) + \frac{\alpha^-}{2}\left(\sqrt{\frac{\beta^-}{\alpha^-}} - 1\right) + \frac{\beta^+}{2}\left(\sqrt{\frac{\alpha^+}{\beta^+}} - 1\right) + \frac{\beta^-}{2}\left(\sqrt{\frac{\alpha^-}{\beta^-}} - 1\right)$$

$$= \sqrt{\alpha^+\beta^+} - \frac{\alpha^+}{2} - \frac{\beta^+}{2} + \sqrt{\alpha^-\beta^-} - \frac{\alpha^-}{2} - \frac{\beta^-}{2}$$

$$= -\frac{1}{2}(\sqrt{\alpha^+} - \sqrt{\beta^+})^2 - \frac{1}{2}(\sqrt{\alpha^-} - \sqrt{\beta^-})^2.$$

This, together with (52), yields

$$\log \mathbb{P}\left(x_i^*(\boldsymbol{Wx}^*)_i \le \gamma \log n\right) \le \frac{1}{2}\left(\gamma' \log\left(\frac{\alpha^+}{\beta^+}\right) - (\sqrt{\alpha^+} - \sqrt{\beta^+})^2 - (\sqrt{\alpha^-} - \sqrt{\beta^-})^2\right)\log n$$

Thus, we have

$$\mathbb{P}\left(x_i^*(\boldsymbol{Wx}^*)_i \le \gamma \log n\right) \le n^{-\frac{1}{2}\left((\sqrt{\alpha^+} - \sqrt{\beta^+})^2 + (\sqrt{\alpha^-} - \sqrt{\beta^-})^2\right) + \frac{\gamma'}{2}\log\left(\frac{\alpha^+}{\beta^+}\right)}.$$

According to the union bound, it holds that

$$\mathbb{P}\left(\min\{x_i^*(\boldsymbol{Wx}^*)_i : i = 1, \ldots, n\} \ge \gamma \log n\right) \ge 1 - n^{1-\frac{1}{2}\left((\sqrt{\alpha^+} - \sqrt{\beta^+})^2 + (\sqrt{\alpha^-} - \sqrt{\beta^-})^2\right) + \frac{\gamma'}{2}\log\left(\frac{\alpha^+}{\beta^+}\right)}.$$

In particular, by $\alpha^+ > \beta^+$ and (5), for sufficiently large n, there exist constants γ and γ' satisfying

$$0 < \gamma + \frac{1}{\log n} < \gamma' < \frac{(\sqrt{\alpha^+} - \sqrt{\beta^+})^2 + (\sqrt{\alpha^-} - \sqrt{\beta^-})^2 - 2}{2\log(\alpha^+/\beta^+)}$$

such that

$$1 - \frac{1}{2}\left((\sqrt{\alpha^+} - \sqrt{\beta^+})^2 + (\sqrt{\alpha^-} - \sqrt{\beta^-})^2\right) + \frac{\gamma'}{2}\log\left(\frac{\alpha^+}{\beta^+}\right) < 0.$$

$\square$

E.2 Proof of Proposition 4 (Contraction Property) and Its Corollary

Proof. We only prove statement (i), then statement (ii) can be shown similarly. According to Lemma 4, Lemma 6, Lemma 7, (16), (38), and (41) hold with probability at least $1 - n^{-\Omega(1)}$ Let $\widetilde{\boldsymbol{\Delta}} := \widetilde{\boldsymbol{A}} - \mathbb{E}[\widetilde{\boldsymbol{A}}]$. According to (42) and (45), we have

$$\boldsymbol{W} = \frac{p^+ - q^+ - \xi(p^- - q^-)}{2}\boldsymbol{x}^*\boldsymbol{x}^{*\top} + \frac{1}{n}(p^+ - \xi p^-)\boldsymbol{E} - (p^+ - \xi p^-)\boldsymbol{I} + \widetilde{\boldsymbol{\Delta}} - (\rho - \mathbb{E}[\rho])\boldsymbol{E}. \tag{53}$$

Let $\boldsymbol{x} \in \mathbb{R}^n$ be arbitrary such that $\|\boldsymbol{x}\| = \sqrt{n}$. Since $\|\boldsymbol{x}\| = \|\boldsymbol{x}^*\| = \sqrt{n}$, we have

$$\|\boldsymbol{x} - \boldsymbol{x}^*\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{x}^*\|^2 - 2\boldsymbol{x}^{*\top}\boldsymbol{x} = 2\boldsymbol{x}^{*\top}\boldsymbol{x}^* - 2\boldsymbol{x}^{*\top}\boldsymbol{x},$$

which implies

$$\left|\boldsymbol{x}^{*T}(\boldsymbol{x}^* - \boldsymbol{x})\right| = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}^*\|^2. \tag{54}$$

Then, we have

$$\|\boldsymbol{Wx} - \boldsymbol{Wx}^*\|$$

$$\le \frac{1}{2}\left|p^+ - q^+ - \xi(p^- - q^-)\right|\|\boldsymbol{x}^*\boldsymbol{x}^{*\top}(\boldsymbol{x} - \boldsymbol{x}^*)\| + \frac{1}{n}\|(p^+ - \xi p^-)\boldsymbol{E}(\boldsymbol{x} - \boldsymbol{x}^*)\|$$

$$\quad + \left|p^+ - \xi p^-\right|\|\boldsymbol{x} - \boldsymbol{x}^*\| + \|\widetilde{\boldsymbol{\Delta}}(\boldsymbol{x} - \boldsymbol{x}^*)\| + \|(\rho - \mathbb{E}[\rho])\boldsymbol{E}(\boldsymbol{x} - \boldsymbol{x}^*)\|$$

$$\le \frac{\sqrt{n}}{4}\left|p^+ - q^+ - \xi(p^- - q^-)\right|\|\boldsymbol{x} - \boldsymbol{x}^*\|^2 + 2\left|p^+ - \xi p^-\right|\|\boldsymbol{x} - \boldsymbol{x}^*\| + \|\widetilde{\boldsymbol{\Delta}}\|\|\boldsymbol{x} - \boldsymbol{x}^*\| + n\left|\rho - \mathbb{E}[\rho]\right|\|\boldsymbol{x} - \boldsymbol{x}^*\|$$

$$\le \frac{c_0 \log n}{4\sqrt{n}}\|\boldsymbol{x} - \boldsymbol{x}^*\|^2 + \left(\frac{2\log n}{n}\left|\alpha^+ - \xi\alpha^-\right| + c_1\sqrt{\log n} + (1 + \xi)\frac{\log n}{\sqrt{n}}\right)\|\boldsymbol{x} - \boldsymbol{x}^*\|, \tag{55}$$

where the first inequality is due to (53) and the triangular inequality, the second inequality follows from (54), and the last inequality follows from (3), (4), Lemma 6 and Lemma 7.

By (16) in Lemma 4, we have

$$\frac{\boldsymbol{W}\boldsymbol{x}^*}{|\boldsymbol{W}\boldsymbol{x}^*|} = \text{sign}(\boldsymbol{W}\boldsymbol{x}^*) = \text{sign}(\boldsymbol{x}^*) = \boldsymbol{x}^*,$$

and

$$\min\{|(\boldsymbol{W}\boldsymbol{x}^*)_i| : i = 1, \ldots, n\} \geq \gamma \log n, \tag{56}$$

Thus, for $\boldsymbol{x}$ satisfying $\|\boldsymbol{x}\| = \sqrt{n}$ and

$$\|\boldsymbol{x} - \boldsymbol{x}^*\| \leq 2c_2 \sqrt{\frac{n}{\log n}} = \frac{24\sqrt{2}c_1}{c_0}\sqrt{\frac{n}{\log n}} \tag{57}$$

we have

$$
\begin{aligned}
\left\|\frac{\boldsymbol{W}\boldsymbol{x}}{|\boldsymbol{W}\boldsymbol{x}|} - \boldsymbol{x}^*\right\| &= \left\|\frac{\boldsymbol{W}\boldsymbol{x}}{|\boldsymbol{W}\boldsymbol{x}|} - \frac{\boldsymbol{W}\boldsymbol{x}^*}{|\boldsymbol{W}\boldsymbol{x}^*|}\right\| \leq \frac{2\|\boldsymbol{W}\boldsymbol{x} - \boldsymbol{W}\boldsymbol{x}^*\|}{\gamma \log n} \\
&\leq \frac{c_0}{2\gamma\sqrt{n}}\|\boldsymbol{x}-\boldsymbol{x}^*\|^2 + \left(\frac{4}{\gamma n}\left|\alpha^+ - \xi\alpha^-\right| + \frac{2c_1}{\gamma\sqrt{\log n}} + \frac{2(1+\xi)}{\gamma\sqrt{n}}\right)\|\boldsymbol{x}-\boldsymbol{x}^*\| \\
&\leq \left(\frac{c_0}{2\gamma\sqrt{n}}\|\boldsymbol{x}-\boldsymbol{x}^*\| + \frac{4|\alpha^+ - \xi\alpha^-| + 2(1+\xi)}{\gamma\sqrt{n}} + \frac{2c_1}{\gamma\sqrt{\log n}}\right)\|\boldsymbol{x}-\boldsymbol{x}^*\| \\
&\leq \left(\frac{12\sqrt{2}c_1}{\gamma\sqrt{\log n}} + \frac{4|\alpha^+ - \xi\alpha^-| + 2(1+\xi)}{\gamma\sqrt{\log n}} + \frac{2c_1}{\gamma\sqrt{\log n}}\right)\|\boldsymbol{x}-\boldsymbol{x}^*\| \\
&= \frac{\left(12\sqrt{2}+2\right)c_1 + 4|\alpha^+ - \xi\alpha^-| + 2(1+\xi)}{\gamma\sqrt{\log n}}\|\boldsymbol{x}-\boldsymbol{x}^*\|,
\end{aligned}
$$

where the first inequality is due to Wang et al. (2020, Lemma 6) and (56), the second inequality follows from (55), and the fourth inequality uses (57). Setting $c_3 = (6\sqrt{2}+1)c_1 + 2|\alpha^+ - \xi\alpha^-| + 1 + \xi$ gives the desired result. $\qquad\square$

Given Proposition 4, we immediately have the following corollary which specifies the linear convergence of the sequence $\{\boldsymbol{x}^t\}_{t\geq 0}$ in the second stage of Algorithm 1 within the contraction region.

Corollary 2. *Suppose that Assumption 1 holds and $\alpha^+, \beta^+, , \beta^-, \alpha^-$ satisfy (5). Let $\{\boldsymbol{x}\}_{t\geq 0}$ be the sequence generated by the second stage of Algorithm 1. Then, for sufficiently large n, either of the following two statements holds with probability at least $1 - n^{-\Omega(1)}$:*

(i) If $\|\boldsymbol{x}^0 - \boldsymbol{x}^*\| \leq \frac{8\sqrt{2}c_1}{c_0}\sqrt{n/\log n}$, then for all $t \geq 0$, we have

$$\|\boldsymbol{x}^t - \boldsymbol{x}^*\| \leq \left(\frac{2c_3}{\gamma\sqrt{\log n}}\right)^t \|\boldsymbol{x}^0 - \boldsymbol{x}^*\|;$$

(ii) If $\|\boldsymbol{x}^0 + \boldsymbol{x}^*\| \leq \frac{8\sqrt{2}c_1}{c_0}\sqrt{n/\log n}$, then for all $t \geq 0$, we have

$$\|\boldsymbol{x}^t + \boldsymbol{x}^*\| \leq \left(\frac{2c_3}{\gamma\sqrt{\log n}}\right)^t \|\boldsymbol{x}^0 + \boldsymbol{x}^*\|,$$

where c_0 and c_1 are the constants in Lemma 3, c_3 is the constant in Proposition 4, and $\gamma > 0$ is the constant in Lemma 4.

E.3 Proof of Proposition 5 (One-Step Convergence)

Proof. We only prove statement (i), then statement (ii) can be shown similarly. Since $\boldsymbol{x} \in \{1, -1\}^n$ and $\|\boldsymbol{x} - \boldsymbol{x}^*\|_2 \leq 2$, then there exists $\ell \in [n]$, such that $\boldsymbol{x} = \boldsymbol{x}^*$ or $\boldsymbol{x} = \boldsymbol{x}^* \pm 2\boldsymbol{e}_\ell$, where $\boldsymbol{e}_\ell$ is an n-dimensional vector with $(\boldsymbol{e}_\ell)_\ell = 1$ and $(\boldsymbol{e}_\ell)_i = 0$ for all $i \neq \ell$. Hence, we consider the following two cases:

1° If $\boldsymbol{x} = \boldsymbol{x}^*$, then by (16) in Lemma 4, it holds with probability at least $1 - n^{-\Omega(1)}$ that

$$x_i^*(\boldsymbol{W}\boldsymbol{x}^*)_i > \gamma \log n > 0,$$

for all $i = 1, \ldots, n$.

2° If $\boldsymbol{x} = \boldsymbol{x}^* \pm 2\boldsymbol{e}_\ell$, then by Lemma 4, it holds with probability at least $1 - n^{-\Omega(1)}$ that,

$$x_i^*(\boldsymbol{W}\boldsymbol{x})_i = x_i^*(\boldsymbol{W}\boldsymbol{x}^* \pm 2\boldsymbol{W}\boldsymbol{e}_\ell)_i = x_i^*(\boldsymbol{W}\boldsymbol{x}^*)_i \pm 2x_i^*(\boldsymbol{W}\boldsymbol{e}_\ell)_i \geq \gamma \log n \pm 2x_i^*(\boldsymbol{W}\boldsymbol{e}_\ell)_i,$$

for all $i = 1, \ldots, n$. Note that $\boldsymbol{W}\boldsymbol{e}_\ell = \left(\widetilde{\boldsymbol{A}} - \rho\boldsymbol{E}\right)\boldsymbol{e}_\ell$, then for $i = 1, \ldots, n$, we have $(\boldsymbol{W}\boldsymbol{e}_\ell)_i = \widetilde{A}_{i\ell} - \rho$. By Lemma 6, we have

$$\left| \widetilde{A}_{i\ell} - (\boldsymbol{W}\boldsymbol{e}_\ell)_i - \frac{p^+ + q^+ - \xi(p^- + q^-)}{2} + \frac{p^+ - \xi p^-}{n} \right| \leq \frac{\log n}{n^{3/2}},$$

which implies that

$$\begin{aligned}
(\boldsymbol{W}\boldsymbol{e}_\ell)_i &\geq \widetilde{A}_{i\ell} - \frac{p^+ + q^+ - \xi(p^- + q^-)}{2} + \frac{p^+ - \xi p^-}{n} - \frac{\log n}{n^{3/2}} \\
&\geq -\xi - \frac{(\alpha^+ + \beta^+ - \xi(\alpha^- + \beta^-))\log n}{n} - \frac{\log n}{n^{3/2}},
\end{aligned}$$

and

$$\begin{aligned}
-(\boldsymbol{W}\boldsymbol{e}_\ell)_i &\geq -\widetilde{A}_{i\ell} + \frac{p^+ + q^+ - \xi(p^- + q^-)}{2} - \frac{p^+ - \xi p^-}{n} + \frac{\log n}{n^{3/2}} \\
&\geq -1 + \frac{(\alpha^+ + \beta^+ - \xi(\alpha^- + \beta^-))\log n}{N} - \frac{p^+ - \xi p^-}{n} + \frac{\log n}{n^{3/2}}
\end{aligned}$$

Therefore, for sufficiently large n, it holds with probability at least $1 - n^{-\Omega(1)}$ that

$$x_i^*(\boldsymbol{W}\boldsymbol{x})_i \geq \gamma \log n \pm 2x_i^*(\boldsymbol{W}\boldsymbol{e}_\ell)_i > 0,$$

for all $i = 1, \ldots, n$.

Combining 1° and 2°, we obtain $x_i^*(\boldsymbol{W}\boldsymbol{x})_i > 0$ for all $i = \{1, \ldots, n\}$, which implies that

$$\frac{\boldsymbol{W}\boldsymbol{x}}{|\boldsymbol{W}\boldsymbol{x}|} = \boldsymbol{x}^*.$$

$\square$

F   Proof of Theorem 1 and Corollary 1

We first prove Theorem 1.

Proof. Suppose with out loss of generality that the following two inequalities hold in Propositions 3 and Corollary 2 simultaneously with probability at least $1 - n^{-\Omega(1)}$:

$$\left\| \boldsymbol{y}^t - \boldsymbol{u}_1 \right\| \leq n\left(\frac{6c_1}{c_0\sqrt{\log n} - 6c_1}\right)^t, \tag{58}$$

$$\left\| \boldsymbol{x}^t - \boldsymbol{x}^* \right\| \leq \left(\frac{2c_3}{\gamma\sqrt{\log n}}\right)^t \left\| \boldsymbol{x}^0 - \boldsymbol{x}^* \right\|, \tag{59}$$

for $\boldsymbol{x}^0$ satisfying $\|\boldsymbol{x}^0\|_2 = \sqrt{n}$ and

$$\left\| \boldsymbol{x}^0 - \boldsymbol{x}^* \right\| \leq \frac{2c_2\sqrt{n}}{\sqrt{\log n}}. \tag{60}$$

Suppose that Algorithm 1 runs T_1 iterations in the first stage and outputs $\boldsymbol{y} \in \mathbb{R}^n$ such that $\|\boldsymbol{y}\|_2 = 1$ and

$$\|\boldsymbol{y}^{(T_1)} - \boldsymbol{u}_1\|_2 \leq \frac{c_2}{\sqrt{\log n}}. \tag{61}$$

By (58), it suffices to let

$$n \left(\frac{6c_1}{c_0\sqrt{\log n} - 6c_1} \right)^{T_1} \leq \frac{c_2}{\sqrt{\log n}},$$

which implies that $T_1 = \mathcal{O}(\log n / \log \log n)$. Then, the second stage of Algorithm 1 starts with $\boldsymbol{x}^0 = \sqrt{n}\boldsymbol{y}^{(T_1)}$, thus by (61) and Lemma 3, we have

$$\|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_2 \leq \|\boldsymbol{x}^0 - \sqrt{n}\boldsymbol{u}_1\|_2 + \|\sqrt{n}\boldsymbol{u}_1 - \boldsymbol{x}^*\|_2 = \sqrt{n}\|\boldsymbol{y}^{(T_1)} - \boldsymbol{u}_1\|_2 + \sqrt{n}\left\|\boldsymbol{u}_1 - \frac{\boldsymbol{x}^*}{\sqrt{n}}\right\|_2$$

$$\leq \frac{c_2\sqrt{n}}{\sqrt{\log n}} + \frac{c_2\sqrt{n}}{\sqrt{\log n}} = 2c_2\sqrt{\frac{n}{\log n}}.$$

Suppose that the second stage of Algorithm 1 terminates at $\boldsymbol{x}^*$ or $-\boldsymbol{x}^*$ with T_2 iterations. By (59), (60), and Proposition 5, we have

$$\left(\frac{2c_3}{\gamma\sqrt{\log n}} \right)^{T_2} \frac{2c_2\sqrt{n}}{\sqrt{\log n}} \leq 2.$$

which implies that $T_2 = \mathcal{O}(\log n / \log \log n)$. $\qquad \square$

To proceed, we require the following lemma.

Lemma 8 (Wang et al. (2020, Lemma 10))**.** *It holds with probability at least $1 - n^{-\Omega(1)}$ that the number of non-zero entries in $\boldsymbol{A}$ is less than $2(\alpha^+ + \alpha^- + \beta^+ + \beta^-)n \log n$.*

Finally, we prove Corollary 1.

Proof. By Theorem 1, the overall iteration complexity is $\mathcal{O}(\log n)$. By Lemma 8, the time complexity of each iteration is $\mathcal{O}(n \log n)$. Hence, the overall time complexity is $\mathcal{O}(n \log^2 n / \log \log n)$. $\qquad \square$